UNIVERSITY
OF SUSSEX

DOCTORAL THESIS

Optimising Triggers and Jet Reconstruction for Higgs Measurement with the ATLAS experiment

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

in the

Experimental Particle Physics Research Group
School of Mathematical and Physical Sciences

Author:
Dan KELSEY

Supervisor:
Dr. Lily ASQUITH

2nd November 2021

ACKNOWLEDGEMENTS

I would never have had the opportunity to undertake this PhD if somebody hadn't taken a chance and believed in me, so the person I would like to thank first is my supervisor, Lily. Your time and wisdom have been invaluable to me and I am incredibly grateful.

I'd like to extend that thank you to the rest of the ATLAS group at Sussex for their combined experience, advice, and the warm and welcoming environment I always found myself in amongst them.

Over the course of my PhD I've been incredibly fortunate to meet many wonderful people at conferences and summer schools. It's always great to bump into a familiar face and catch up over a beer. I'd like to especially thank all those who I spent many a happy hour with at CERN and exploring Geneva, France and Switzerland. Those are adventures and memories I'll cherish for a lifetime.

Brighton itself has been an amazing adventure and I've made so many dear friends. In particular, the Abacrew, who made my first year so amazing, and the Hove Homies, who kept me sane for all the following! The dinner parties, pub-sessions, board game nights and general letting-off-of-steam made me feel so at home. Thank you all.

I'd like to thank my friends from the University of Kent who kept me going through my undergrad. I'd also like to specially mention both Jamie, without whom I would have been very lonely, and Anthony, without whom I would have been very productive.

Thank you to my friends from home, for sticking with me all these years: Liam, Ryan, James, George and Jake. I've learned more from you all than I could ever say and I'm so lucky to have friends full of such care, curiosity and character.

A very important mention goes to my family. My Mum and Dad, who have always worked so hard to give me the wonderful life with which I am privileged, and who have always loved and supported me. My Nan, Grandma, Grandpa and my late Grandad, who have always shown me I can do so much, and never stopped believing in me. And Aaron, who only comes last because he is the youngest; I'm incredibly proud to have a brother like you, and I hope I can make you proud too. Thank you all once again for never stifling my thirst for learning, despite me being an insufferable know-it-all!

Most of all, I'd like to thank my dearest Anna. Without you, I would never have finished this. Your love and belief in me has kept me strong and kept me going, all whilst doing a PhD of your own. You're an inspiration to me.

I, Dan KELSEY, hereby declare that this thesis has not been and will not be, submitted in whole or in part to another university for the award of any other degree.

Brighton,
2nd November 2021

Dan KELSEY

University of Sussex
School of Mathematical and Physical Sciences
Experimental Particle Physics Research Group

DAN KELSEY, DOCTOR OF PHILOSOPHY

Optimising Triggers and Jet Reconstruction for Higgs Measurement with the ATLAS experiment

---

ABSTRACT

Since observation of the Higgs boson, one of the fundamental priorities of the ATLAS experiment is to precisely measure properties of this particle, allowing evaluation of the Standard Model (SM) of particle physics. Associated Higgs production with a top-quark pair ($t\bar{t}H$) is particularly useful as it can directly determine the absolute value of the top-Yukawa coupling.

Correctly identifying events containing interactions with the Higgs is challenging: among all $t\bar{t}H$ final states, that in which a Higgs decays to a $b\bar{b}$-pair and and the top quarks each decay hadronically to jets has the largest branching ratio, but also the least signal purity. The overwhelming majority of backgrounds within ATLAS come from non-resonant production of multijet events. The other significant background in this search comes from $t\bar{t}$+ jets events.

Distinguishing signal events from background events based around the identification and classification of jets becomes increasingly difficult as pile-up increases; proposed increases in luminosity at the LHC will make this worse. Optimising the ATLAS trigger and jet reconstruction are required to combat this effect. In the trigger, I investigate the viability of tracking with jets, opening new avenues to improving identification.

Developments in jet reconstruction, through access to jet substructure (JSS), can improve flavour-tagging. Limitations arise due to contributions from Non-Global Logarithms (NGL), a problem circumnavigated by utilising soft drop grooming. By analysing JSS variables in flavour-tagged soft drop-groomed jets I am able to identify optimal grooming strengths and variables for discriminating between different jet flavours. In particular, I find that applying soft-drop grooming to tau-flavoured jets improves correlation between the jet mass and the mass of tau leptons. Comparing reconstructed jets with truth jets at different grooming strengths also allows me to evaluate the validity of reconstructed jets in representing real physics. Finally, I use all this to compare results from Monte Carlo (MC) simulation of $t\bar{t}$, $t\bar{t}H$ and dijet events, as well as real data from the ATLAS detector in order to optimise discrimination between Higgs events and background. These processes display potential for optimising Higgs measurements with flavour-tagging improvements through JSS analysis in soft drop-groomed small-jets.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

To me there has never been a higher source of earthly honour or distinction than that connected with advances in science.

Isaac Newton

In 2012 the universe changed dramatically - or at least our fundamental understanding of it did - with the discovery of the Higgs boson. This discovery by the ATLAS and CMS collaborations at the Large Hadron Collider (LHC) was the final puzzle-piece of the Standard Model (SM): the theoretical framework that lies at the heart of particle physics [24][25]. First formulated during the 1960s and 1970s, the SM explains the foundational principles of the universe in terms of elementary particles, their interactions, and the universe-shaping forces that emerge. The Higgs boson was the last of these particles to be experimentally verified, and its discovery opens the door to new regions of scientific exploration. The SM has its limitations, and there exists phenomena within nature it cannot explain. Rigorous analysis of the Higgs boson, its properties and interactions, allows us to evaluate proposed models vying to supersede the SM as the keystone of particle physics. There also exists the possibility that a comprehensive understanding of this momentous particle will lead us to a new understanding of physics.

A particular focus is the Yukawa coupling between the Higgs and fermions, leading to the masses of the latter. The fermion masses are proportional to the strength of their coupling with the Higgs; therefore, understanding the coupling of the heaviest of these, the top quark, is especially important. A direct measurement of the Yukawa coupling between the Higgs and top quarks is accessible via the $t\bar{t}H$ Higgs production channel, known as associated production with top quarks.

Jets, collimated sprays of hadronically decaying particles, are of particular importance in decoding the physics of Higgs production. The Higgs decay channel with the highest branching ratio is Higgs decay to a b-quark pair ($H \rightarrow b\bar{b}$), resulting in significant hadronic output. Signal purity in this decay channel is poor: the dominant process at the LHC is multi-jet production, and the dominant background for $H \rightarrow b\bar{b}$ is top pair-production ($t\bar{t}$) with additional

hadronic jets. Understanding the physics of hadronic jets is a fundamental part of decoding Higgs physics at ATLAS; however, proposed future changes at the LHC to increase luminosity pose growing issues with jet reconstruction due to the increasing effect of pile-up. Optimising jet reconstruction to improve accuracy, reliability and efficiency is, therefore, vital to analysis.

Developments within the ATLAS Jet Signature trigger to improve jet identification at high speed is one aspect of this, leading to testing possible solutions, such as hardware implementations like the Fast TracKer (FTK). Advances in offline analysis rely on exploring innovative techniques and processes, including the use of JSS and novel grooming techniques in jet reconstruction. To address issues posed by to traditional jet grooming techniques, such as NGL, alternatives grooming techniques must be explored that overcome this, such as soft drop grooming.

The main analysis of this thesis focuses on the application of soft drop grooming to small Anti-k_T jets constructed from calorimeter clusters. I inspect the changes on JSS observables brought about by soft drop grooming in a number of different jet flavours. This is done using Monte Carlo (MC) samples of $t\bar{t}$, $t\bar{t}H$ and dijet events, as well as real data.

This thesis is presented in nine chapters. The first, Chapter 1, discusses the theoretical fundamentals of the SM and, in particular, the Higgs boson. Following this, the experimental setup of the LHC and the ATLAS detector are introduced in Chapter 2. Chapter 3 explains the generation of MC samples used to model physics events at ATLAS, as well as the algorithms used for reconstructing physics objects using signals from the detector subsystems. Discussion on the methods of jet identification, reconstruction and processing can be found in Chapter 4. In Chapter 5, I consider the ATLAS hardware and software triggers and potential tracking developments in the ATLAS HLT Jet trigger. The main analysis of the thesis, relating to the effects of soft drop grooming on flavour-tagged small jets, is presented in Chapter 6. This is followed by a final conclusion with recommendation for future analysis.

THE STANDARD MODEL AND THE HIGGS BOSON

1

These things thou must always have in mind: What is the nature of the universe, and what is mine?

Marcus Aurelius

In this chapter, I will introduce the theoretical structure that underpins the particle physics work within this thesis. I shall start this chapter with the most successful theory in particle physics, the Standard Model (SM). The SM provides us with a theoretical framework that describes the universe in the form of elementary particles and how their interactions generate the fundamental forces of nature. An overview of these particles and forces are given in Section 1.1.

The achievements of the SM are evidenced by numerous experimental successes, predicting results to a high precision. One recent (and well-publicised) success involves the process of electroweak symmetry breaking, described by the Brout-Englert-Higgs mechanism in 1967, and the discovery in 2012 of a new particle, the mechanism's proposed scalar boson: the Higgs boson[24, 25]. These are covered in Sections 1.2 and 1.3.

1.1 The Standard Model

The SM of particle physics is the theoretical foundation of particle physics and describes the fundamental physical phenomena at microscopic levels of nature. It explains vast amounts of experimental observations made within particle physics, alongside some cosmological phenomena from the early universe. Various sections were developed over the course of the 20th century, a combination of experimental results and theoretical formalisations, culminating in the 1960s and 70s [26, 27]. Described as one of the most successful scientific theories in history, it has been rigorously tested by experiment, and has made a number of correct predictions [28], including many measureable quantities which have been well verified experimentally. It describes the nature and interactions of three out of the four fundamental forces of nature: the electromagnetic (EM) force, the weak force and the strong force. The final force, gravity, is not

explained within the SM; however, it is much weaker than the other three forces – its effects are too small to affect physics at the small scales of particle physics.

The SM is a Quantum Field Theory (QFT) and integrates the foundational principles of special relativity and quantum mechanics [29]. It describes the phenomenology of fundamental fields and their interactions, and identifies particles as excitations within their corresponding quantum fields, based within a four-dimensional Minkowski spacetime. The SM describes these particles as point-like, meaning they contain no internal structure and, therefore, are the base components of the universe. It is an effective theory, and gives a description of these particles and their effects via the exchanges of gauge bosons.

Described within the model are three generations of fermions, four gauge vector bosons and one scalar boson, as well as a number of qualities they possess. The SM has predicted the existence of multiple particles that have since been experimentally discovered: the top quark [30, 31], the τ lepton [32] and the W [33] and Z bosons [34], all in the last century, and in 2012 the experimental confirmation of the SM was complete with discovery of the Higgs boson[24, 25].

The SM is the most general renormalisable QFT that is also invariant under transformations of gauge symmetries. It is therefore an example of a gauge theory, whereby the Lagrangian is invariant under a continuous group of local transformations.

Gauge theories and symmetries within the Standard Model

Symmetries and conservation laws are key to the SM. In 1915, mathematician and theoretical physicist Emmy Nöether demonstrated that symmetries within a system in nature are indicators that the system possesses conserved physical properties. Specifically, she showed that each differentiable symmetry of the action of these systems has a corresponding conservation law [35]. The action of a physical system can be defined as the integral over time of a Lagrangian density function ($\mathcal{L}$):

$$S = \int \mathcal{L} dt \tag{1.1}$$

This makes use of the Lagrangian formalism to describe the field dynamics of the system. All necessary field dynamics are described within the $\mathcal{L}$ functions, where the total is split into a kinematic and interaction term:

$$\mathcal{L}_T = \mathcal{L}_{free} + \mathcal{L}_{int} \tag{1.2}$$

This describes both the spacetime evolution of the free field, and the way in which these different fields are coupled to one another. The Lagrangian of the SM can be described as the combination of the Lagrangians describing the Electroweak (EW) force, Quantum Chromodynamics (QCD) and the Higgs field:

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_{Higgs} \tag{1.3}$$

These symmetries described can be either observable, or simply intrinsic to the system. Under certain transformations, these symmetries allow the properties of the system to be maintained. This is gauge theory, a specific class of quantum field theories.

Within gauge theories exists a group of transitions of the field variables, known as gauge transformations. These gauge transformations must leave the basic physics of the quantum field unchanged. The transformations form mathematical groups, allowing us to use group theory to identify and classify symmetries conserved within the SM. Within gauge theories the Lagrangian is invariant under a continuous group of local transformations. Gauge invariance, as this is known, bestows the chosen theory with a certain symmetry. Within the SM this is realised as invariance of fermion fields under given transformation T_X, where X denotes the conserved property of the field [36].

The apparatus of the group of gauge transformations within a given gauge theory necessitates limitations on the properties of the fields it describes, such as interactions with other fields and particles. Consequently, the forces of nature are described as gauge fields, and interactions are delivered through exchanges of the quanta of the field, gauge bosons, within the gauge field of the analogous local symmetry group. The collection of the gauge transformations, the group of the SM, is the Lie Group, containing all transformations between possible gauges [37]. The associated Lie algebra of group generators is associated to any Lie Group. For every group generator there emerges a corresponding gauge field.

The symmetry group of the SM can be mathematically described as

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{1.4}$$

Here, the first part of the group, $SU(3)_C$, denotes the theory of Quantum Chromodynamics (QCD), providing a description of the strong nuclear force. $SU(2)_L \otimes U(1)_Y$ refers to EW Theory, a combined theory of Quantum Electrodynamics (QED), describing the force of electromagnetism, and the weak nuclear force [38]. However, this symmetry is spontaneously broken and leaves both forces behaving differently from one another at everyday energies. Here the subscripts refer to the charge of the fields described by the symmetry: C signifies the colour charge, L denotes left-handedness and Y indicates the weak-hypercharge. Weak hypercharge relates to the correspondence between the electric charge of electromagnetism, Q, and weak isospin. This leaves electromagnetism with a $U(1)_{EM}$ symmetry group.

The SM can be described as a Non-Abelian Yang Mills QFT [39]. Here an Abelian Group is one with commutativity – whereby the order of group operations when applying to two elements of the group, does not affect the result. Non-Abelian Lie groups, however, behave such that the generators for gauge fields do not commute; therefore, this means that the gauge bosons described within these theories can self-interact. Yang-Mills theory involves use of the special unitary group, $SU(n)$, using non-Abelian Lie groups. Both the weak and strong forces are described by these non-Abelian groups.

QED, the first QFT proposed, is an Abelian gauge theory, with the $U(1)_{EM}$ symmetry group. It has a single gauge field, the electromagnetic four-potential, and has the photon as its single gauge boson. The unified EW theory is also a Yang-Mills theory.

1.1.1 Particles

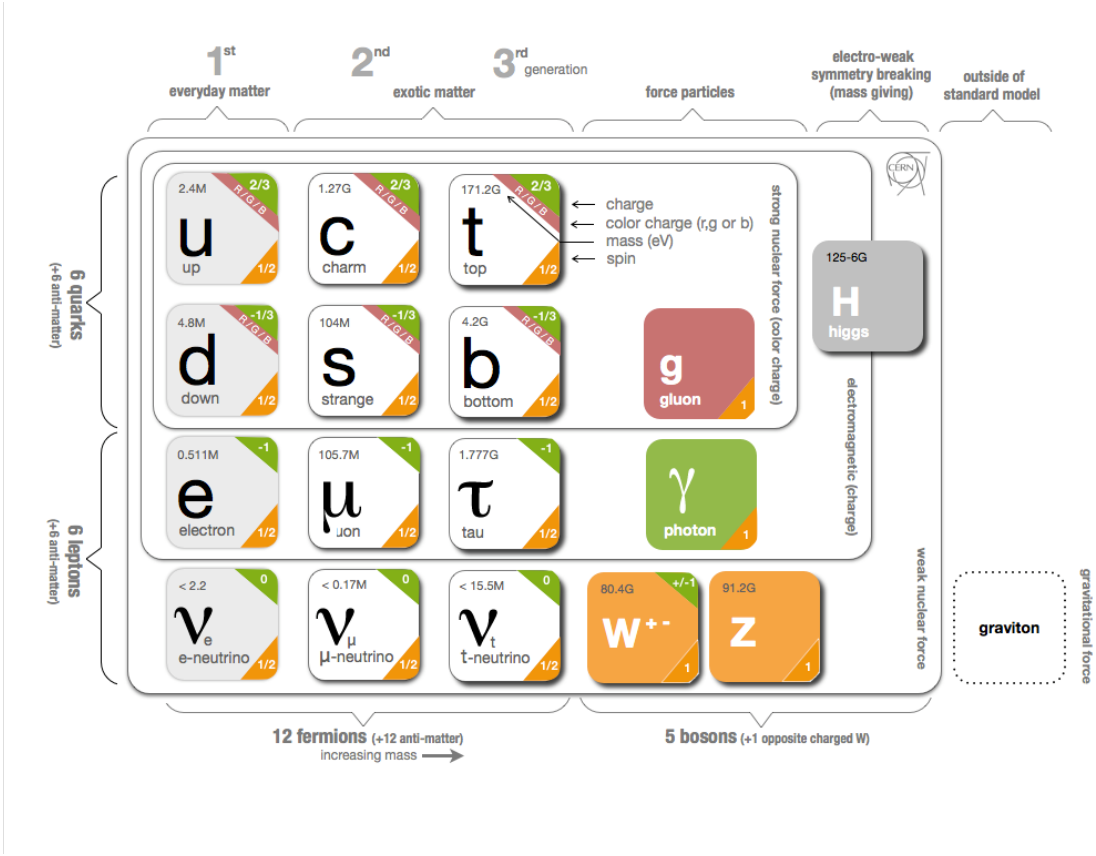


Figure 1.1: Summary of the elementary particles in the SM, including thier respective electric charge, colour charge, mass and spin [1].

The SM also defines the set of fundamental particles that form the basic building blocks of nature, see Figure 1.1. These particles are excitations in their corresponding, underlying quantum fields. We can divide these particles into two broad groups: fermions and bosons [37]. Fermions have a $\frac{1}{2}$ integer spin in units of $\hbar$ and obey the physics of Fermi-Dirac statistics. Bosons, described as force-propagating particles, have integer values of spin, once more in units of $\hbar$.

Beyond spin, there are three other fundamental properties particles possess; mass, helicity and chirality. Mass determines the propagation of particles though spacetime in the absence of interactions. Helicity is the projection of a particle's spin onto its momentum and is defined as "left-handed" when opposed to the momentum of the particle, or "right-handed" if in the same direction. Massive particles have a helicity that can change, depending on the inertial reference frame. Chirality is a related concept, with no physical representation and is equivalent to the helicity in the case of massless particles. It is an intrinsic quantum property describing the wavefunction of the particle.

Fermions

Fermions, further split into leptons and quarks, can be described as the "matter particles": those which construct the basis of matter. Fermions have a $\frac{1}{2}$ integer spin in units of $\hbar$ and obey the physics of Fermi-Dirac statistics [40]. Leptons only couple with the weak and electromagnetic forces, but quarks couple with the weak, electromagnetic and strong force (through Quantum Chromodynamics (QCD)).

We can further subdivide leptons and quarks into 3 generations of doublets. The first generation of both leptons and quarks collectively form the basis for stable matter: the electron (e), up quark (u) and down quark (d). Amongst the quarks and the charged leptons, each generation has successively higher mass than the last, and mass varies by several orders of magnitude. This difference in mass is not built into the mathematical foundations of the SM, and therefore is a free parameter of the theory. Leptons come in six flavours, subdivided into electrically charged leptons and electrically neutral neutrinos.

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \qquad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \qquad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix} \tag{1.5}$$

The electron (e), muon (μ) and tau (τ) leptons have a charge of -1 in units of the electron charge, e. The flavours of the three neutrinos reflect the flavours of the three charged leptons within each generations: the electron neutrino (ν_e), muon neutrino (μ_e) and tau neutrino (τ_e). Having no electric charge means the neutrinos only couple to the weak force. Leptons possess a value known as a lepton number, L, related to their flavour. Both electrons and electron neutrinos have an electron number value, L_e, of -1. The values of muon and τ generation leptons relate to the muon number, L_μ, and τ number, L_τ, respectively. Each lepton number is conserved.

Quarks come in six flavours in total. These are organised into 3 generations of "up-type"/"down-type" pairs. The "up-type" quarks are, in ascending order of mass, the up (u), charm (c) and top (t) quarks. Each of these three quarks possess an electric charge of $+\frac{2}{3}$ in units of the electron charge, e. The down (d), strange (s) and bottom (b) quarks are the "down-type" quarks and each carry an electric charge of $-\frac{1}{3}$ in units of the electron charge, e. These can be organised into three $SU(2)$ doublets constructed from the "up-type" and "down-type" quark within each generation.

$$\begin{pmatrix} u \\ d \end{pmatrix}, \qquad \begin{pmatrix} c \\ s \end{pmatrix}, \qquad \begin{pmatrix} t \\ b \end{pmatrix} \tag{1.6}$$

Each quark also carries a baryon number of $\frac{1}{3}$. In nature, quarks are not found in isolation: they form bound states called hadrons with integer values of electric charge and Baryon number. Hadrons can be categorised further into mesons and baryons, according to the number of quarks (and anti-quarks) they are made of. Mesons are made of an even number of quarks (and anti-quarks), such as the pion (π^0), made of an up quark-antiquark pair ($u\bar{u}$). Baryons consist

of an odd number of quarks (and anti-quarks), such as the proton (uud) and neutron (udd). The Baryon number is always conserved. Quarks are also colour charged, meaning that they couple with the strong force. The charges of these quarks are (anti-)red, green and blue.

For all Fermions there exists an anti-particle partner, as predicted by the Dirac equation, with identical mass and spin, but with opposite charges [41]. This includes electric charge, colour charge, Baryon number and Lepton number.

Bosons

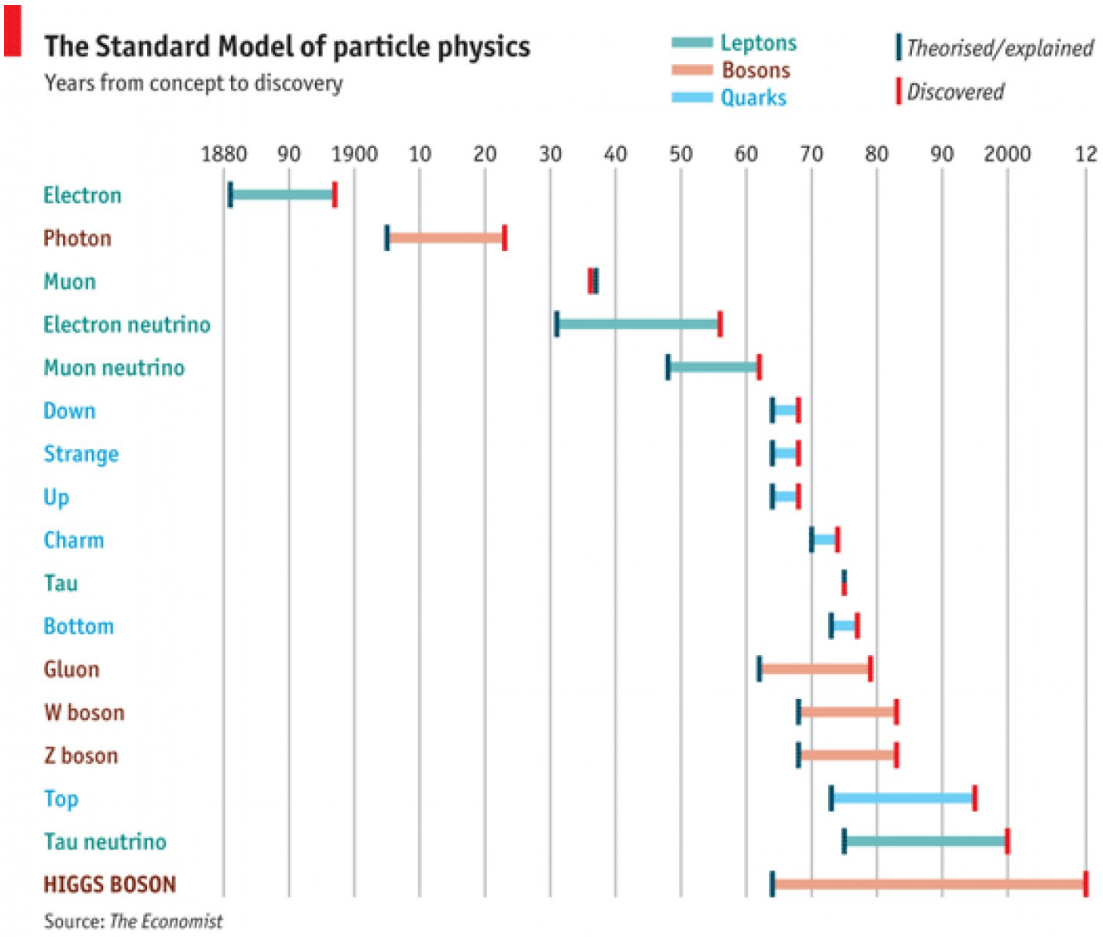


Figure 1.2: Timeline of the theoretical predictions and experimental observation of elementary particles of the SM [2]

Bosons, force-mediating particles, have integer values of spin, once more in units of $\hbar$, and observe Bose-Einstein statistics [40]. There are four bosons described within the SM. Three gauge bosons, or vector bosons, all with a spin value of 1, and a single scalar boson (the Higgs boson) with a spin of 0. The first of the vector bosons is the photon, γ, the mediator of the electromagnetic (EM) force. This force couples with particles with an electric charge, Q, measured in units of the electron charge, e.

$$e = 1.602 \times 10^{19}\,\mathrm{C} \tag{1.7}$$

Two of the three remaining vector bosons are the mediators of the weak force: the two $W^{\pm}$ bosons and the Z^0 bosons. The final of the vector bosons is the gluon, g, the exchange particle of the strong force. Gluons are self-interacting, meaning they feel the effects of the strong force themselves. The only other particles they couple with are the quarks. Both quarks and gluons are colour charged: quarks come in three and gluons come in eight colour charges. Beyond the SM, a mediator of gravity, the spin 2 graviton, has been proposed, but not experimentally confirmed [42].

1.1.2 Forces and Interactions

The SM contains within it a comprehensive description of three of the four fundamental forces of nature. These forces can be described in terms of couplings between particles and fields, from which the particles of the SM arise. The dynamics of these fields can be explained through use of the Lagrangian field density, $\mathcal{L}$.

$$\mathcal{L} = \mathcal{L}(\psi, \partial_\mu \psi) \tag{1.8}$$

Here, ψ is a fermion field and $\partial_\mu \psi$ is the partial derivative with respect to the four-vector of all generalised space coordinates, x^μ. The invariance of the Lagrangian density function under particular transformations relates to symmetries within the system described. A simple example of one such transformation is a phase transformation:

$$\psi \rightarrow e^{-i\theta}\psi \tag{1.9}$$

The SM requires these symmetries be local, gauge symmetries, relying on a continuous change in phase that is dependent on spacial coordinates such that

$$\psi(x) \rightarrow e^{-i\theta(x)}\psi(x) \tag{1.10}$$

Under this transformation extra terms will be introduced due to the partial derivative acting on the local dependence. To maintain the gauge invariance of the $\mathcal{L}$, the partial derivative must be replaced with a covariant derivative:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iA_\mu \tag{1.11}$$

Here we see the introduction of a new vector field, A_μ. The addition of this vector field ensures the extra terms introduced are dealt with such that the $\mathcal{L}$ remains invariant:

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu \tag{1.12}$$

The inclusion of this new vector field has the consequence of introducing excitations within this field: the force-carrying bosons. The interaction between these bosons and the fermion field, ψ, is also inserted. For each example of this, whereby a field describing a force is developed into our $\mathcal{L}$, we must also introduce a coupling constant, α, to determine the interaction strength between our fermion field and the bosons. This coupling is dependent on the energy scale of our system, and this is known as a "running coupling".

Our full $\mathcal{L}$ will include the description of the EW force, QCD and the effects of the Higgs field, as shown in equation 1.3.

1.1.3 Electromagnetic Interaction

The first fundamental force to be described through the use of QFT is the EM force, in the theory of Quantum Electrodynamics (QED) [29]. QED arises from the $U(1)_{Y\omega}$ symmetry group [43]. This group has only one generator, meaning the theory predicts there will be a single boson mediating this force. The symmetry of the system leads to QED being invariant under gauge transformations of the global phase transformation type, such that

$$\psi \rightarrow \psi' = e^{(-iQ\theta)}\psi \tag{1.13}$$

Here ψ is our fermion field, of spin $\frac{1}{2}$ and Q is electric charge. The particle described meets the requirements for the Dirac equation of motion such that

$$i\gamma^{\mu}\partial_{\mu}\psi - m\psi = 0 \tag{1.14}$$

Where γ^{μ} represents the set of Dirac gamma matrices. To ensure invariance under gauge transformations, we set our covariant derivative with the requirement

$$\partial_{\mu} \rightarrow D_{\mu} \equiv \partial_{\mu} - iQA_{\mu} \tag{1.15}$$

The introduction of a new vector field, A_{μ} reveals a massless, spin-1 gauge boson, the mediator of the EM force: the photon (γ). Therefore, QED describes the coupling between charged fermions and the photon. QED is an Abelian group of symmetry $U(1)_Q$. The commutativity of the 1-dimensional group results in the photon being electrically neutral, and therefore unable to self-interact. The $\mathcal{L}$ for QED is given by

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \psi\left(i\gamma^{\mu}D_{\mu} - m\right)\psi \tag{1.16}$$

Here we have defined the EM field tensor, $F_{\mu\nu}$ as

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} \tag{1.17}$$

The coupling constant for the EM interaction defined at zero momentum transfer is

$$\alpha_{EM} = \frac{e^2}{4\pi} \approx \frac{1}{137} \tag{1.18}$$

Where e is the elementary charge. The EM force is considered a chiral gauge group, as it couples with right- and left-handed particles at different strengths.

1.1.4 Weak Interaction

The second force described within the SM is the weak force. Experimental results have revealed that the weak force only interacts with left-handed particles. By only acting on left-handed components of the fermion fields, and only coupling with left-handed neutrinos, the weak force violates parity conservation [44]. Weak interactions couple to the weak isopin, T, related to the chirality of particles. Weak interactions conserve T_Z. Fermions with left-handed chirality

have a T_Z value of $\frac{1}{2}$ for up-type quarks and neutrinos, and T_Z value of $-\frac{1}{2}$ down-type quarks and charged leptons. These left-handed fermions can be described as a two-component field and grouped into doublets of isospin:

$$\psi_l = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \qquad \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \qquad \begin{pmatrix} t_L \\ b_L \end{pmatrix} \tag{1.19}$$

$$\psi_l = \begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}, \qquad \begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}, \qquad \begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix} \tag{1.20}$$

These doublets transform under the symmetry group $SU(2)_L$. The Lagrangian of weak force is

$$\mathcal{L}_W = i\psi_l \gamma^\mu D_\mu \psi_l + i\bar{e_R}\gamma^\mu \partial_\mu e_R \tag{1.21}$$

This is invariant under the $SU(2)_L$ transformation

$$\psi_l \rightarrow {\psi_l}' = e^{\frac{1}{2}\vec{\alpha}(x)\vec{\tau}}\psi_l \tag{1.22}$$

And the Covariant derivative is

$$D_\mu = \partial_\mu + i\frac{g}{2}\vec{\tau}\overrightarrow{W_\mu}(x) \tag{1.23}$$

$\vec{\tau}$ represents the Pauli matrices and g is coupling constant of the weak force. The weak force is derived from a non-Abelian $SU(2)_L$ symmetry and contains three generators, and therefore three gauge bosons mediating interactions. These are the two $W^\pm$ bosons and the Z boson. The lack of commutativity means that these bosons interact. The subscripted L in the symmetry group references the weak force coupling only to particles with left-handed chirality (or anti-particles with right-handed chirality). The weak force is composed of both charged and neutral currents mediated by these bosons.

Imposing local gauge invariance results in three component vector fields:

$$W_\mu^{i=1,2,3} \tag{1.24}$$

Mixing between these three gauge fields gives rise to the physical W boson. From W_μ^1 and W_μ^2 we can form a linear construction to produce the $W^\pm$ boson:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp W_\mu^2\right), \tag{1.25}$$

The $W^\pm$ bosons carry an electric charge $Q = \pm 1$ and only interacts with left-handed fermions through the weak isospin charge, T_Z. The $W^\pm$ boson has a T_Z value of ± 1. The transmission of the W boson transforms particles weakly from an isospin value of $T_Z = +\frac{1}{2}$ to $T_Z = -\frac{1}{2}$. Right-handed fermions have a weak isospin value $T_Z = 0$, and therefore do not couple with the W boson. Consequently, they do not undergo weak transformations, remaining as singlets.

$$u_R, d_R, c_R, s_R, t_R, b_R, e_R, \nu_{e,R}, \mu_R, \nu_{\mu,R}, \tau_R, \nu_{\tau,R} \tag{1.26}$$

The Z boson is electrically neutral. It couples to both left- and right-handed fermion fields, but in a different manner: for left-handed particles the coupling depends on both the electric

charge, Q, and the weak isospin element T_Z, for right-handed particles the coupling with the Z boson is proportional to the electric charge. Imposing that only left-handed fermions are in doublets means that interactions involve W_μ^3 and left-handed fermions, but not right-handed. The Z boson, therefore, is not trivially W_μ^3, meaning Z and W are not mass degenerate, as would be the case if W_μ^3 and Z were same. The Z boson has a weak isospin component $T_Z = 0$ and couples to both right- and left-handed fermions.

Right handed neutrinos, referred to as sterile neutrinos, have not been observed in nature. They would not couple with the weak force, and therefore would hypothetically only interact with gravity. Neutrinos in the SM are only presented as left-handed fermions, with right-handed anti-particles.

The charged weak-current interacts with quarks in such a way that their flavour is changed. This is the only coupling that can change the flavour of quarks, and leads to parity violation within the SM. The probability of these transformation are outlined within the Cabibbo– Kobayashi– Maskawa (CKM) matrix [45, 46]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.27}$$

This relates the mass eigenstates of quarks to their flavour eigenstates (denoted d'). The diagonal elements, relating to the transition within generations, are close to 1, meaning that these transformations are most likely.The weak force is shown to have the smallest range of the fundamental forces, acting over just 10×10^{-18} m. This indicates that the $W^\pm$ and Z bosons are short-lived, and are therefore have large masses.

1.1.5 Electroweak Interaction

The weak and EM interactions explain physics in the universe below the $\mathcal{O}(100\,\text{GeV})$; however, above this limit they fail to do so. Therefore, theories containing concealed gauge-invariance obscured by symmetry-breaking at low energies were developed to tackle this issue, leading to the unification of the weak force and EM first demonstrated by Glashow, Weinberg and Salam [26, 27, 47, 48, 49]. As a result, the search began for an unseen gauge variance, only observable above threshold energies that lead to a theory containing symmetry breaking.

The result of this symmetry breaking is that at low energies the only perceived gauge invariance is that of QED, and the two appear as separate interactions. This ultimately lead to the development of unified force of weak interactions and EM - Electroweak (EW) theory - a theory that incorporates elements of QED and the Fermi theory of weak interactions.

The hidden symmetry group of this interaction is

$$SU(2)_L \otimes U(1)_{Y\omega} \tag{1.28}$$

Here L indicates the coupling of the weak force to left-handed components of fermion fields. The identification Y represents the weak hypercharge, a union of T_Z, the projection of isospin in the z-direction, and electric charge, Q, defined in the Gell Mann-Nishijima equation [50, 51]

$$Y = 2(Q - T_Z) \tag{1.29}$$

The $\mathcal{L}$ of the EW interaction is

$$\mathcal{L}_{EW} = i\psi_l \gamma^\mu D_\mu \psi_l + i\bar{e}_R \gamma^\mu \partial_\mu e_R \tag{1.30}$$

This includes the covariant derivative

$$\partial_\mu \rightarrow D_\mu \equiv \partial_\mu - ig\vec{\tau} \cdot \overrightarrow{W_\mu} - ig'\frac{Y}{2}B_\mu \tag{1.31}$$

W_μ and B_μ are the gauge fields of the two symmetry groups, $SU(2)_L$ and $U(1)_{Y\omega}$. g and g' stand for the coupling constants of each symmetry respectively. We impose the requirement that this $\mathcal{L}$ is invariant, and therefore must institute a triple gauge field. This gauge field contains three generators, to be described as

$$W_\mu^{i=1,2,3} \tag{1.32}$$

$SU(2)_L$ therefore has three component gauge bosons, the bosons that are associated with weak force interactions. $U(1)_{Y\omega}$ has a single massless generator, the photon (γ), therefore this replaces the $U(1)_Q$ symmetry from QED. This gives a total of four gauge bosons. These three W bosons will transmit the weak force, and the B_μ boson established above interacts with the weak hypercharge, $Y\omega$ and therefore mediates the EM force. These four non-physical gauge bosons are all predicted by the theory to be massless. They are related to the physical bosons that transmit electroweak (ew) interactions: $W^\pm$, Z and γ. From W_μ^1 and W_μ^2 we can form a linear construction to produce the $W^\pm$ boson:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp W_\mu^2\right) \tag{1.33}$$

The Z boson and photon vector fields arise from mixing from W_μ^3 and B_μ in such a way that the photon fields do not couple with neutrinos:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{1.34}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W, \tag{1.35}$$

$$A_\mu = {B_\mu}^3 \cos\theta_W + W_\mu^3 \sin\theta_W \tag{1.36}$$

Where the field A_μ once more represents the photon. The photon, Z and W bosons appear as combinations of the eigenstates of the symmetry group. $\sin\theta_W$ is the Weinberg, or weak mixing angle [52]. This is defined as a ratio of the two ew coupling constants such that:

$$\tan\theta_W = \frac{g'}{g} \tag{1.37}$$

This value has been determined experimentally.

The $W^{\pm}$ bosons carry an electric charge $Q = \pm 1$ and only interacts with left-handed fermions through the weak isospin charge, T_Z. The Z boson is electrically neutral. It couples to both left- and right-handed fermion fields, but in a different manner. For left-handed particles the coupling depends on both the electric charge, Q, and the weak isospin element T_Z. For right-handed particles the coupling with the Z boson is proportional to the electric charge, Q. The Lagrangian accordingly becomes

$$\mathcal{L}_{EW} = \psi_l \gamma^\mu \left(\partial_\mu - i g \overrightarrow{\tau} \overrightarrow{W_\mu} - i g' \frac{Y}{2} B_\mu \right) \psi_l \bar{e_R} \gamma^\mu \left(\partial_\mu - i g' \frac{Y}{2} B_\mu \right) e_R - \frac{1}{4} \overrightarrow{W_{\mu\nu}} \overrightarrow{W^{\mu\nu}} - \frac{1}{4} \overrightarrow{B_{\mu\nu}} \overrightarrow{B^{\mu\nu}} \quad (1.38)$$

where

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.39)$$

and

$$W_{\mu\nu} = \partial_\mu \overrightarrow{W_\nu} - \partial_\nu \overrightarrow{W_\mu} + i g \overrightarrow{W_\mu} \times \overrightarrow{W_\nu} \quad (1.40)$$

The first two terms in the Lagrangian relate to the kinetic energy of the fermions and their couplings with the gauge fields. The final two terms describe the energy of the gauge fields and their self-interactions. The Lagrangian is invariant under local gauge the transformations

$$\psi_l \rightarrow {\psi_l}' = e^{i\left(\vec{\alpha}(x)\overrightarrow{\tau} + \beta(x)\frac{Y}{2}\right)} \psi_l \quad (1.41)$$

$$e_R \rightarrow {e_R}' = e^{\beta(x)\frac{Y}{2}} \psi_l \quad (1.42)$$

The weak force is shown to have the smallest range of the fundamental forces, acting over just 10×10^{-18} m. This indicates that the $W^{\pm}$ and Z bosons are short-lived, and are therefore massive. This contradicts our Lagrangian which requires that the gauge bosons are massless, otherwise invariance is broken. Nevertheless, the mass of the W and Z bosons have been experimentally confirmed [38]. Theories of the weak force that incorporate massive gauge bosons lead to problems with renomalisation or loss of gauge invariance. This leaves a requirement for an apparatus allowing both preservation of gauge symmetry in the SM and to solve the discrepancy between the mass differences between verified physical particles and those gauge bosons described in theory. Therefore developments to the symmetry breaking in EW theory that could explain this were devised. The particular structure of the spontaneous symmetry breaking of this unified theory leads to the development of a complex scalar field, the Higgs field, which couples to fermions and bosons, giving them their mass [53, 54]. See Section 1.2.

1.1.6 Strong Interaction

The third and final of the fundamental forces described by the SM is the strong force, described by Quantum Chromodynamics (QCD). It arises from the symmetry group $SU(3)_C$, where the charge, C, stands for "colour" [55, 56]. One property of this symmetry is that it is a non-chiral

gauge group and therefore acts the same on right- and left-handed particles. We also impose invariance under local transformations. We therefore have to construct eight new three-by-three generators describing interactions of eight gluon fields, $G_\mu^{a=1,\dots 8}$, to maintain the local gauge invariance of the symmetry. The covariant derivative is given here,

$$\partial_\mu \rightarrow D_\mu \equiv \partial_\mu - i g_s T_a G_\mu^a \tag{1.43}$$

Here g_s is the strong coupling constant, also written as

$$\alpha_s = \frac{g_s^2}{4\pi} \tag{1.44}$$

G_μ^a represents the gluon fields, with α_s taking values 1 to 8. The symmetry group has eight generators, therefore predicts eight bosons. T_a represents the eight $SU(3)_C$ generators, and uses the Gell-Mann matrix, λ_a

$$T_a = \frac{1}{2}\lambda_a \tag{1.45}$$

QCD defines the interactions between these eight gluon fields, and the only particles that interact with the colour-charge: quarks and gluons. Gluons are colour-charged, they are non-Abelian and, therefore, self-interacting. Gluons generate virtual gluons in numbers that are proportional to the distance between interacting colour-charged particles.

The strong force is aptly named, as it is the strongest of all the fundamental forces. At zero-momentum transfers the coupling constant, $\alpha_S = 1$. At these low energies QCD cannot be described through perturbation theory; however, the running coupling of the strong force is highly sensitive to the energy scale due to asymptotic freedom [57, 58]. This means that the coupling constant of the strong force at high energies becomes small enough for perturbation theory to be used. It becomes so small, in fact, that at very high energies quarks can be approximated as free particles. The strong coupling constant, α_S is a running coupling that is dependent on the separations between particles interacting. This is due to the gluon self-interaction loop processes.

Two critical consequences of the asymptotic nature of the running coupling are Asymptotic Freedom and Colour Confinement. Asymptotic Freedom is the uniform growth in the coupling strength between gluons and quarks as the energy scale of interactions decreases, and as the distance between these interacting particles increases. The running coupling can be described as diverging asymptotically at large distance, or low energy. Therefore the strong force coupling between colour charged particles grows higher as the distance between them increases. The result of Asymptotic Freedom is that inside hadrons, where quarks can be considered to be in bound states, quarks and gluons interact very little and roam about “near free”. At high energies quarks can be modelled as free particles, and perturbative calculations can therefore make very accurate predictions.

The reverse effect of this is Colour Confinement. Colour-charged particles - quarks and gluons - cannot exist in isolation: they must always form colour neutral hadrons, either as a triplet within a baryon or a doublet within a meson. When the quark components of a hadron are pulled apart, such as a quark-antiquark pair within a meson, the strong force interaction

between them increases. Eventually the energy required to overcome the strong force interaction and maintain the distance between the quark-antiquark pair becomes so high that it is instead more energetically favourable for an additional quark anti-quark pair to be created to maintain the colour confinement of the diverging quarks. This process is called Hadronisation. Hadronisation will repeat until the energy of the colour charged particles is low enough for bound states of colour-neutral hadrons to form. The cascading effect of these hadronising particles leaves a tell-tale signature within particle physics experiments in the form of jets. The consequence of this is that isolated colour charged particles, such as quarks, are unable to be observed before they can return to states of colourless hadrons. The one deviation from this rule is the top quark, (t), which is so massive that it decays before Hadronisation can occur. The SM establishes the mean lifetime of the top quark on the $\mathcal{O}(5\times10^{-25}\,\text{s})$, about twenty times shorter that the time needed for strong interactions to take place. This allows the unusual opportunity to study an isolated quark. Due to the effects of Colour Confinement, the range of the strong force is almost as small as that of the weak force, on the $\mathcal{O}(10^{-15}\,\text{m})$, even though gluons are massless. Quarks can be defined as colour triplets

$$q=\begin{pmatrix} q_r \\ q_g \\ q_b \end{pmatrix} \tag{1.46}$$

The principle of colour-charge also allows quarks of the same flavour to exist in the same bound state without violating the Pauli Exclusion Principle, as long as they have differing colour charges. An example is the baryon Δ^{++}, made of three up quarks, uuu. Any states with a net colour-charge are not invariant under $SU(3)_C$ transformations – those which are colour neutral are invariant. These neutral states can be formed from three quarks (or anti quarks) of the different (anti-)colour charges, red, $green$ and $blue$, in a baryon, from quark-antiquark pairs where colour charges cancel, or hadrons formed of combinations of these baryons and mesons.

The development of the Coupling Constant through the change of energy scale can be determined from the Lagrangian of QCD and the masses of the quarks [59]. This can be defined as a function of the energy transfer squared:

$$\alpha_S\left(\mu^2\right)=\frac{12\pi}{\left(33\text{–}2n_f\right)\ln\left(\frac{\mu^2}{\Lambda^2{}_{QCD}}\right)} \tag{1.47}$$

Here $\Lambda_{QCD}\approx 200\,\text{MeV}$ and is a scale of QCD theory and n_f is the number of quarks with a mass below the transferred energy. The Coupling Constant decreases as a function of this scale. Coupling diverges rapidly when μ decreases and approaches this scale. This means that the theory is only perturbative at high energy scales, as $\alpha_S<1$ is only true for $\mu\gg\Lambda_{QCD}$.

$$\mathcal{L}=-\frac{1}{4}\sum_{a=1}^{8}G^a_{\mu\nu}G^{a\mu\nu}+\sum_{f=1}^{6}\left[\bar{q}_f i\gamma^\mu\left(\partial_\mu+igG_\mu\right)q_f-m_f\bar{q}_f q_f\right] \tag{1.48}$$

Here

$$G_\mu = \frac{1}{2}\sum_{a=1}^{8} G_\mu^a \lambda^a \tag{1.49}$$

This is a sum over each of the eight gluon states. λ^a are the Gell-Mann matrices. f refers to the different quark flavours, and q_f refers to the quark triplet state, shown in Equation 1.46.

1.2 The Higgs Mechanism

EW theory and QCD do well to describe the nature of fundamental particles and their interactions, and their combined use matches experimental results with great accuracy; however, the foundational propositions of chiral symmetry and gauge invariance seemingly do not allow an explanation for the mass of bosons that we observe. The development of the SM from QFT representations of the fundamental forces allows the formation of a description of nature where massless fermion fields interact with massless force-mediating gauge bosons. Both the photon, γ, and the gluon, g, mediate long range forces, and are therefore massless. This means they do not contradict this prediction. Nevertheless, from experimental results we know that the W and Z bosons are massive, as are the fermions, leaving the SM incomplete. The expansion of EW theory to include the origin of these masses lead to a number of proposals.

The Lagrangian predicts SM particles to be massless, and difficulties exist when attempting to reconcile this. The total Lagrangian of the SM can be modelled as the sum of each of the forces within the SM describing particle interactions, $\mathcal{L}_{EW}$ and $\mathcal{L}_{QCD}$, with the masses of all fundamental particles:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{Mass} \tag{1.50}$$

Inserting the mass terms by hand is problematic, as it voids the gauge invariance and leads to divergences within the SM. The global Lagrangian symmetry must be preserved, even as the gauge symmetry is broken, to ensure the theory is renormalisable. The way this is fixed is by introducing spontaneous symmetry breaking. This lead to independent proposals by Englert, Brout, and Higgs that an element of the broken symmetry within the ew theory could encompass massive elementary particles within the SM, and in particular could explain the masses of the W and Z bosons mass [53, 54]. Of these, only Higgs propsed that this this involve a new, massive boson.

The proposal starts with an electrically neutral scalar complex field, named the Higgs field. The interaction between this field and other particles gives these particles mass. The Higgs field couples to both fermion and boson fields, and is a complex scalar doublet which will transform under the group $SU(2)_L$ as a $T_Z = \frac{1}{2}$ doublet. It has four real components and models mass generation of these fundamental particles through the mechanism of spontaneous symmetry

breaking:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.51}$$

Where ϕ^+ and ϕ^0 are generic complex fields with the structure

$$\phi^+ = \frac{1}{\sqrt{2}}\left(\phi_1 + i\phi_2\right) \tag{1.52}$$

$$\phi^0 = \frac{1}{\sqrt{2}}\left(\phi_3 + i\phi_4\right) \tag{1.53}$$

Each consists of complex scalar field, with two degrees of freedom. ϕ_1 and ϕ_2 represent the real and imaginary parameterisations of the Higgs potential, $V\left(\phi_1, \phi_2\right)$. The Lagrangian can be written as

$$\mathcal{L} = \left(\partial_\mu \phi\right)^* \left(\partial^\mu \phi\right) - V\left(\phi\right) \tag{1.54}$$

Where $V\left(\phi\right)$ represents the potential

$$V\left(\phi\right) = \mu^2 \left|\phi\right|^2 + \lambda \left|\phi\right|^4 \tag{1.55}$$

To ensure this scalar field is renormalisable and invariant under the symmetry group of the EW theory

$$SU(2)_L \otimes U(1)_{Y\omega} \tag{1.56}$$

We ensure the Higgs potential takes the following structure

$$V\left(\phi\right) = \mu^2 \phi^\dagger \phi + \lambda\left(\phi^\dagger \phi\right)^2 \tag{1.57}$$

The first term of the scalar potential, parameterised by μ, represents the scalar mass term. The second term represents the self-interaction vertex and involves the parameter λ. This potential is symmetric under rotations through ϕ space. We require that the potential energy is bounded from below, $\lambda < 0$; however, we can choose the parameter μ freely. If we assume the constants μ and λ to both be real, this gives the Higgs potential a parabolic shape, with a single minimum at $\phi^0 = 0$. Although, if we choose μ to be imaginary, and therefore $\mu^2 < 0$, we obtain a distinctive shape with a continuum of minima that can help explain spontaneous symmetry breaking. This is named the "Mexican Hat" potential, See Figure 1.3.

This time ϕ^0 simply represents a local minimum. The continuum of these minima as positions about the circular trough of the 'hat' means that there is no single global minima. The shape of the potential means that the field is forced to choose one of these arbitrary minima, and this breaks the rotational gauge symmetry. The convention of representation is

$$\langle \phi_3 \rangle = v \tag{1.58}$$

$$\langle \phi_1 \rangle = \langle \phi_2 \rangle = \langle \phi_4 \rangle = 0 \tag{1.59}$$

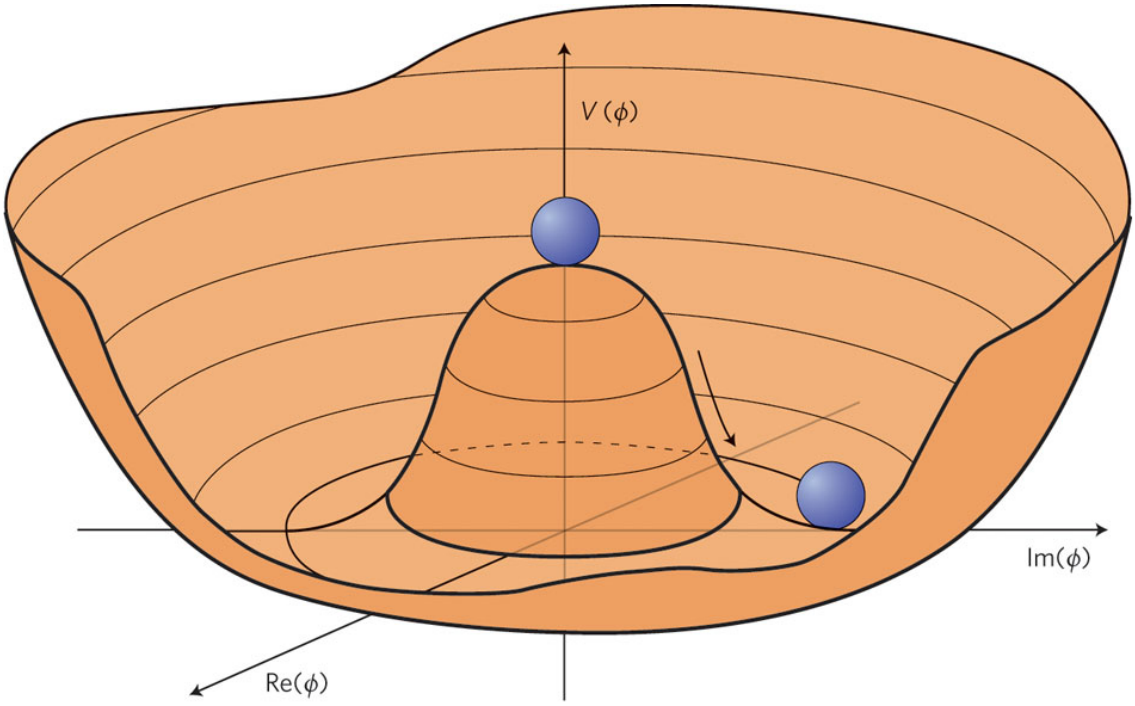


Figure 1.3: Visualisation of the Higgs potential, also known as the "Mexican Hat" potential, in the complex plane. The lowest-energy state corresponds to a randomly chosen point within the "trough". [3].

v is known as the Vacuum Expectation Value (VEV), v, and can be written as

$$v \equiv \sqrt{\frac{-\mu^2}{\lambda}} \tag{1.60}$$

It is non-zero and has the value

$$v = \sqrt{\frac{1}{\sqrt{2}G_F}} \approx 246\,\text{GeV} \tag{1.61}$$

Hhere G_F is the Fermi constant. The radius of the circle of minima is defined as

$$|\phi_0| = \pm\sqrt{\frac{v^2}{2}} = \pm\sqrt{\frac{-\mu^2}{2\lambda}} \tag{1.62}$$

This allows us to write a perturbation of the Higgs field, H about the minimum.

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{1.63}$$

At high energies the symmetry remains unbroken and the field will take the central VEV, corresponding to $V(0,0)$. As the energy of the system passes below a threshold, the field minimises and selects a random non-zero ground state from one of the infinite possible minima, the Vacuum Expectation Value (VEV), within the trough of the Mexican Hat. This VEV of the Higgs field is not gauge-invariant, therefore the gauge-symmetry is spontaneously broken in the ground state of the vacuum. The Higgs field can be expanded about the chosen minimum with additional scalar fields, each chosen with a zero-value VEV. These are η and ξ.

$$\phi = \frac{1}{\sqrt{2}}\Big((\eta + v) + i\xi\Big) \tag{1.64}$$

The expanded Lagrangian is now

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \eta\right)\left(\partial^\mu \eta\right) - \frac{1}{2}\left(\partial_\mu \xi\right)\left(\partial^\mu \xi\right) - \frac{1}{2}m_\eta^2 + const + \mathcal{O}\left(\eta, \xi\right) \tag{1.65}$$

The two terms in this equation both describe bosons: $\eta + v$ represents excitations in the η field and ξ represents excitations in the direction of the field that maintains the same potential. Including this parameterisation within the Lagrangian of the Higgs gives access to terms describing the kinetic energy and interactions of these two fields, and that of the gauge. It also allows us to specify a term describing the coupling between the gauge field and η. We can see a mass term for the η field and mass terms for the EW bosons are also introduced: $W^{i=1,2,3}$ and B_μ. The higher-order terms in η and ξ represent the field interactions. η represents the Higgs field, its excitations are a massive boson, the Higgs boson. Of the four additional degrees of freedom established in the Higgs mechanism, three emerge as "Goldstone" Bosons [60]. For each continuous symmetry broken a massless scalar particle manifests. ξ relates to one of these spinless, massless "Goldstone" scalar bosons – although, these decidedly cannot exist in nature - the inability for a gauge boson to transform into a scalar boson prevents it.

The Goldstone bosons are represented within the complex scalar doublet as a complex scalar field $G^\pm$ and a real scalar field G^0:

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} G^+ \\ \frac{1}{\sqrt{2}}\left(v + H + iG^0\right) \end{pmatrix} \tag{1.66}$$

The real scalar field H is the Higgs boson field. This can be removed through use of the correct gauge transformation, though, leading to the creation of a massive gauge boson with an extra longitudinal polarisation component. The Goldstone bosons created from the symmetry breaking are absorbed by the weak gauge field, and these become the massive W and Z bosons.

The mass of the $W^\pm$ boson is determined through a combination of the VEV of the Higgs field, v, with the coupling constant of the $SU(2)_L$ gauge interaction, g:

$$M_W = \frac{gv}{2} \tag{1.67}$$

The masses of the Z boson and photon comes from mixing of W_μ^3 and B_μ, and involve the two coupling constants of the EW theory, g and g'.The Z boson becomes massive, and the photon maintains no mass:

$$M_Z = \frac{\left(g^2 + g'^2\right)v}{2} \tag{1.68}$$

$$M_\gamma = 0 \tag{1.69}$$

The masses of the W and Z bosons are also related by the Weinberg angle in a way that is consistent with experimental measurements:

$$M_W = \frac{M_Z}{\cos\left(\theta_W\right)} \tag{1.70}$$

The fourth and final degree of freedom from the Higgs field forms a massive, spinless boson, called the Higgs boson. The mass of the Higgs boson is given as

$$M_H = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2} \tag{1.71}$$

As the parameter λ is not given by theory the mass of the Higgs boson is a free-parameter of the theory, and therefore can only be found through experiment.

Whilst this explains the origin of mass for the weak gauge bosons, we must go further to discover the reason for massive fermions, expanding the full Lagrangian of the Higgs field allows us to explore its coupling to fermion fields. Investigation of the simplest form of a mass term shows it is not gauge invariant:

$$m\bar{\psi}\psi = m(\bar{\psi_L}\psi_R + \bar{\psi_R}\psi_L) \tag{1.72}$$

This is because, as proposed in EW theory, the right-handed components of fermions are singlets, whereas left-handed elements are described by doublets. This means that they will transform differently, breaking invariance. To overcome this, the Higgs field must be introduced in the following fashion

$$\mathcal{L}_f = -Y_f\left(\bar{\psi_L}\phi\psi_R + \bar{\psi_R}\phi^\dagger\psi_L\right) \tag{1.73}$$

Here we present the Yukawa Coupling, Y_f, as an interaction between a scalar and Dirac type field (the fermion doublet). The subscript f relates to the applicability of all fermions. The gauge invariance will now be preserved due to the use of the $SU(2)$ doublets. We can demonstrate how this works for the case of lepton doublets

$$\begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix} \tag{1.74}$$

Inserting this into our newly formed Lagrangian gives

$$\mathcal{L}_\ell = Y_\ell\left[\left(\bar{\nu_\ell}, \bar{\ell}\right)_L \begin{pmatrix} 0 \\ v+H \end{pmatrix} \ell_R + \bar{\ell_R}(0, v+H)\begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L\right] = -\frac{Y_f v}{\sqrt{2}}\bar{\ell}\ell - \frac{Y_f H}{\sqrt{2}}\bar{\ell}\ell = -m_\ell\bar{\ell}\ell - \frac{m_\ell}{v}H\bar{\ell}\ell \tag{1.75}$$

The two new terms that we encounter symbolise the mass of the lepton, m_ℓ, and the coupling they have with the Higgs boson. An expansion of this can also be used to pinpoint the origin of quark masses in the SM. The relation between these masses and the Yukawa coupling is

$$m_f = Y_f\frac{v}{\sqrt{2}} \tag{1.76}$$

The nature of the Yukawa coupling varies for each fermion field, and therefore the mass can be unique for each particle. This mechanism allows fermions to become massive without nullifying the gauge-invariance of weak interactions. However, the coupling, and therefore these masses, are not predicted by theory, so they must once again be confirmed by experiment. We can consider these as free parameters of the SM, and measurements of the Yukawa coupling enable rigorous tests of this theory.

1.3 The Higgs Boson

Following the development of the Higgs Mechanism to explain the origin of masses in the SM the final piece of the puzzle was the search for the predicted Higgs boson: a spinless, electrically neutral gauge boson, and the last component of predicted fundamental particles within the SM. On 4th July 2012 a joint announcement by the collaborations of ATLAS and Compact Muon Solenoid (CMS) at CERN (the configuration of which will be discussed in Chapter 2) declared the independent discoveries of the Higgs boson [24, 25].

First proposed in 1964, the search for the Higgs boson since its prediction in theory is the longest of any fundamental particle [53, 54]. Its discovery not only confirmed the validity of the Higgs mechanism to explain the origin of mass for electroweak gauge bosons, it opened a new avenue for particle physics to test the accuracy of theoretical predictions through precision measurements of the properties Higgs boson - for example, the Higgs boson is the first scalar elementary particle discovered in nature - and its interactions with the other fundamental particles. The mass of the Higgs boson is one property requiring experimental data, as it is a free parameter of the SM, and therefore not predicted in theory. The mass of the Higgs boson has been measured as (125.10 ± 0.14) GeV [28].

One way to scrutinise the nature of this discovered particle is through analysis of the various predicted and observed ways the Higgs boson can be produced within nature, and the particles into which it will decay. In-depth examination of the frequencies and likelihoods of these productions and decays can fortify our understanding of the SM, accurately testing theoretical predictions relating to the coupling of the Higgs boson and the other fundamental particles of nature. The production and decay of the Higgs boson has been observed and measured in multiple channels since its discovery and the data is so far highly compatible with the predicted Yukawa coupling between the Higgs boson and other particles of the SM. Over several orders of magnitude theoretical predictions for these Yukawa couplings strongly agree with experimental evidence.

1.3.1 Higgs Boson Production and Decay

Production of the Higgs boson within particle experiments can take place through a number of different production channels. There are four dominant production processes of the Higgs at the centre-of-mass energy of the LHC, $\sqrt{s} = 13\,\text{TeV}$. The most prevalent is gluon-gluon fusion (ggF), followed by vector boson fusion (VBH), associated vector boson production (VH) and associated top-quark pair production ($t\bar{t}H$) [28]. Both the VH and VBH production allow analysis of the coupling between the Higgs and the weak gauge bosons, W and Z. The ggF and $t\bar{t}H$ processes permit investigation of direct coupling between the Higgs boson and the quarks, enabling measurement of the Yukawa couplings between these particles. Although both the ggF and $t\bar{t}H$ production-processes allow direct examination of the Higgs coupling to the top quark, the ggF process is close to two orders of magnitude more prevalent as a produc-

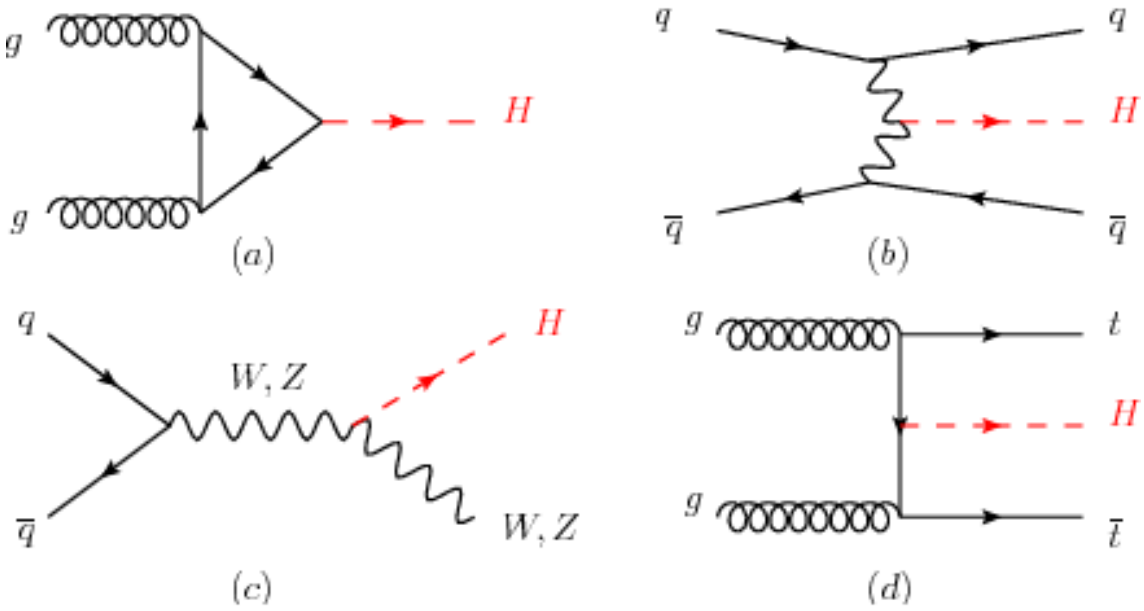


Figure 1.4: Feynman diagrams of SM Higgs production in (a) gluon fusion, (b) weak-boson fusion, (c) Higgs-strahlung (or associated production with a gauge boson) and (d) associated production with top quarks [4].

tion mode when compared to $t\bar{t}H$. Problematically, though, the gluons in this scenario couple to the Higgs boson through a virtual quark loop. This assumes no Beyond Standard Model (BSM) processes taking place within the virtual quark exchange loop, meaning that accurately studying this production method is made more difficult. Conversely, the $t\bar{t}H$ production method relies only on tree-level computations of Higgs coupling, therefore the model dependence of the measurement is reduced when using this production channel to examine the coupling.

The lifetime of the Higgs is 10^{-22} s, therefore it decays before it can be detected within experiment. Information about the Higgs boson must, therefore, be constructed through observed properties of its decay products, of which there are a number. Discovery of the Higgs in 2012 relied on utilising a number of these channels, especially the decay into a pair of Z bosons ($H \to Z^*$) and decay into a pair of photons ($H \to \gamma\gamma$) [24, 25]. Following the discovery of the Higgs, searches for evidence of other predicted decays have led to the observation of the first fermionic decays. In 2016 the Higgs decaying to a pair of τ leptons ($H \to \tau\tau$) was observed [61]. As can be seen in Fig 1.6 the most common decay process for a Higgs of mass 125 GeV is the decay to a b-quark and b-antiquark ($b\bar{b}$) pair. This decay channel, ($H \to b\bar{b}$), was only recently observed in 2018 [62, 63].

The Yukawa coupling between the Higgs boson and fermions is directly proportional to the mass of the fermion, therefore it is strongest with the top quark, the heaviest fermion with a mass of (173.0 ± 0.4) GeV [28]:

$$Y_t = \frac{\sqrt{2}m_t}{v} \approx 1 \tag{1.77}$$

Finally discovered in 1995, the top quark was the final quark predicted by the SM to be observed [30, 31]. The lifetime of the top quark is even shorter, on the scale of 10^{-25} s. As mentioned, this means that it is exceptional amongst quarks in that it decays before it is able to hadronise.

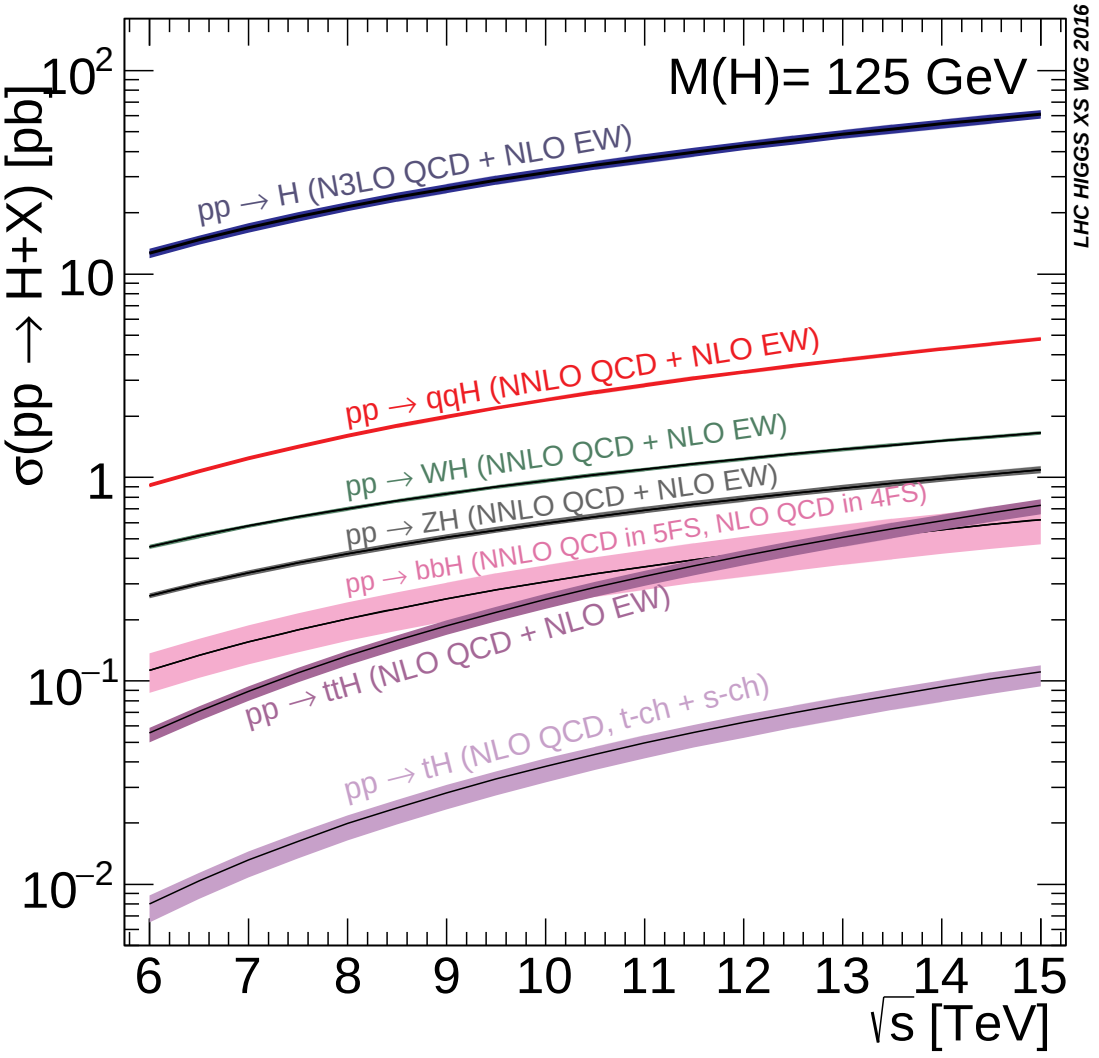


Figure 1.5: The cross-sections for selected production modes for the SM Higgs, of mass 125 GeV, and their uncertainties as a function of the CoM energy of the LHC [5]

The top quark is also distinctive for being heavier than the W boson – it is therefore the only quark that decays semi-weakly, producing a real W boson. The SM predicts that all top quark decays result in a W boson and a down-type quark (d, s, b). The Cabibbo–Kobayashi–Maskawa (CKM) matrix (Equation 1.27) determines the likelihood of quark flavour-changing. The values describing mixing between the top quark and the down quark, $|V_{td}|$, and the mixing between the top quark and the strange quark, $|V_{ts}|$, are found to contribute less than 5%:

$$|V_{td}| \approx 8 \times 10^{-3} \tag{1.78}$$

$$|V_{ts}| \approx 35 \times 10^{-3} \tag{1.79}$$

The mixing between the top quark and the bottom quark, $|V_{tb}|$, is dominant. In practise this means that the top quark decaying into a W boson and b quark is the only notable decay process.

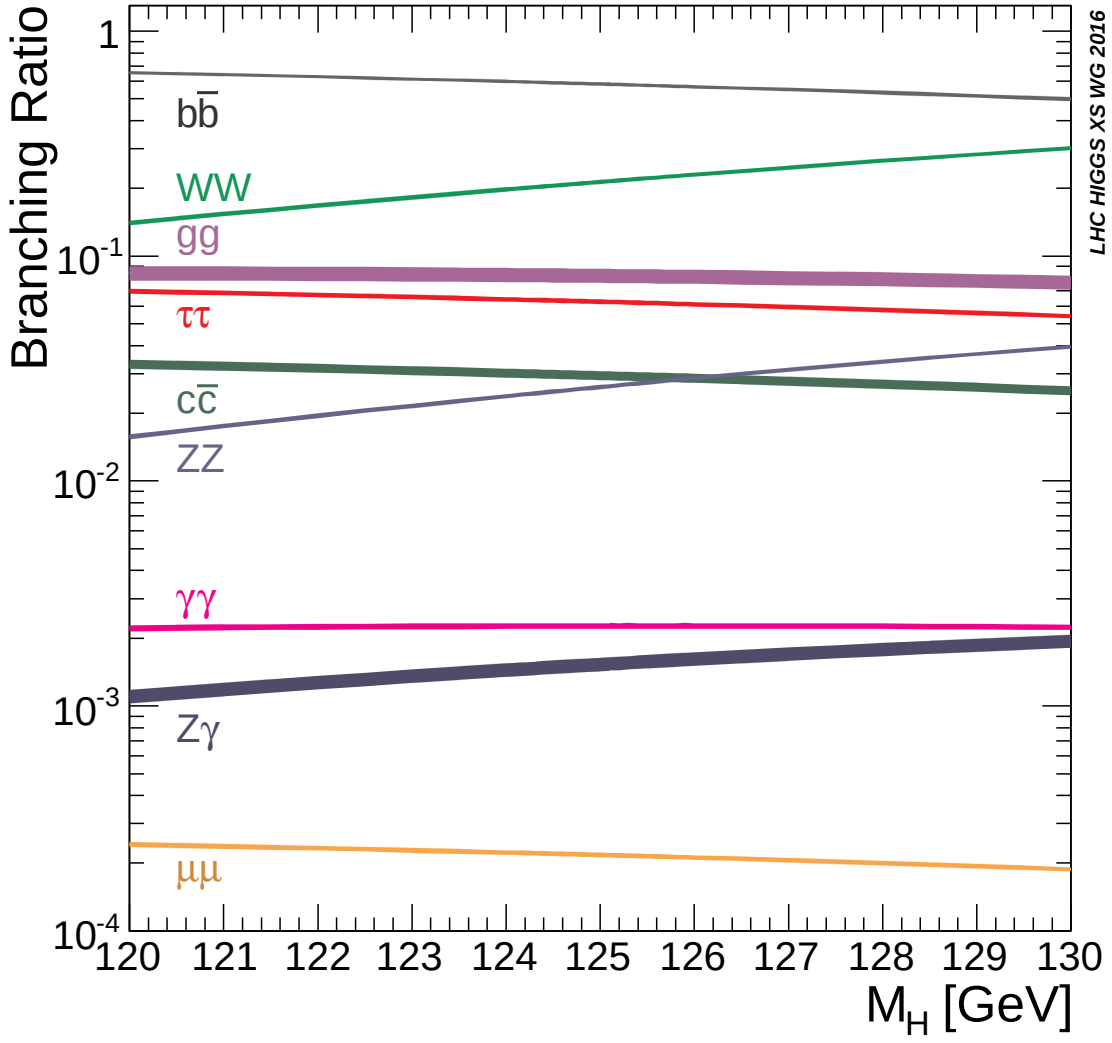


Figure 1.6: Branching Ratios and their uncertainties for the decay processes of the SM Higgs boson as a function of mass [5]

The decay of a $t\bar{t}$ pair can be classified by the subsequent decay of the W bosons that are produced, see Figure 1.8. We find the most likely process is that both of these W bosons decay hadronically, into quark-antiquark pairs ($q\bar{q}$) [28]. As the quarks in this final state evolve into hadronic jets, this process is known as the all-hadronic, or all-jet final state. Following this the second most likely process is the decay of one W boson hadronically, and one leptonically, such that it decays into a lepton and a neutrino ($\ell^- \nu_\ell$) . Least likely is a dilepton decay ($\ell\ell$) of both W bosons. Their respective contributions are given below.

$$t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow q\bar{q} b q\bar{q} \bar{b} \quad (45.7\%) \tag{1.80}$$

$$t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow q\bar{q} b \ell^- \bar{\nu}_\ell \bar{b} + \ell^+ \nu_\ell b q\bar{q} \bar{b} \quad (43.8\%) \tag{1.81}$$

$$t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow \ell^+ \nu_\ell b \ell^- \bar{\nu}_\ell \bar{b} \quad (10.5\%) \tag{1.82}$$

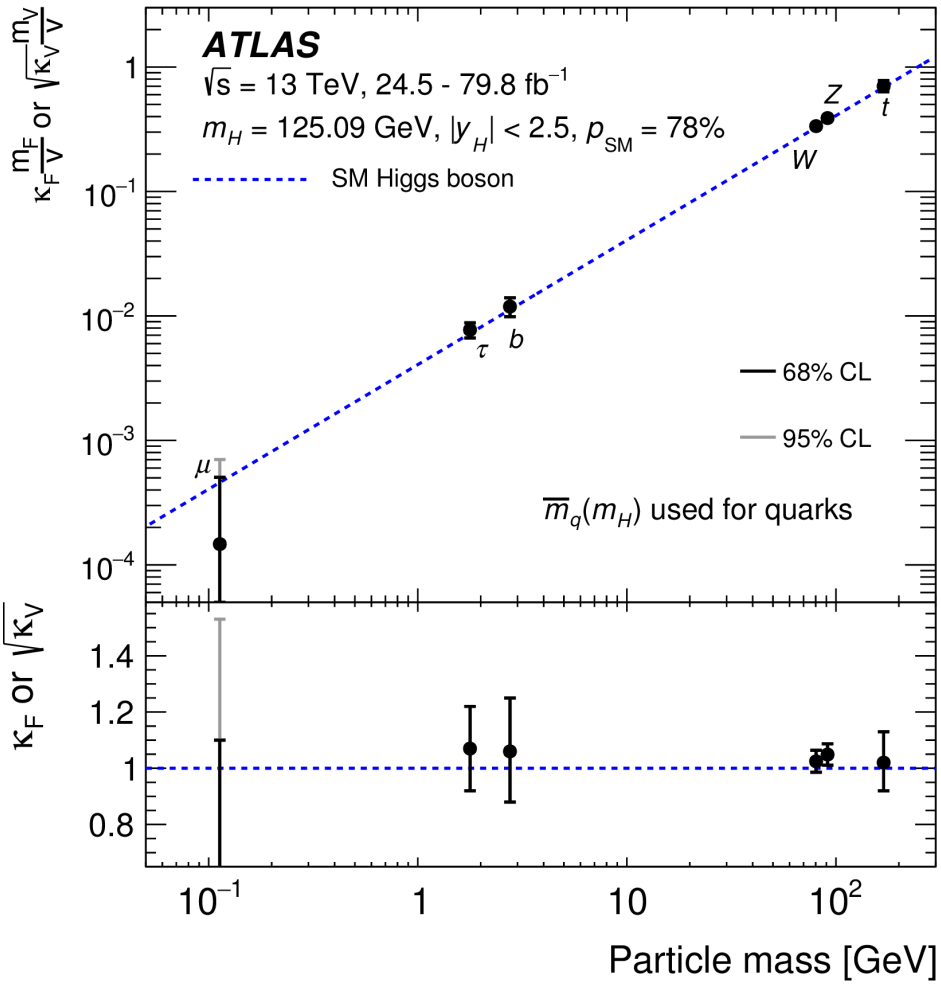
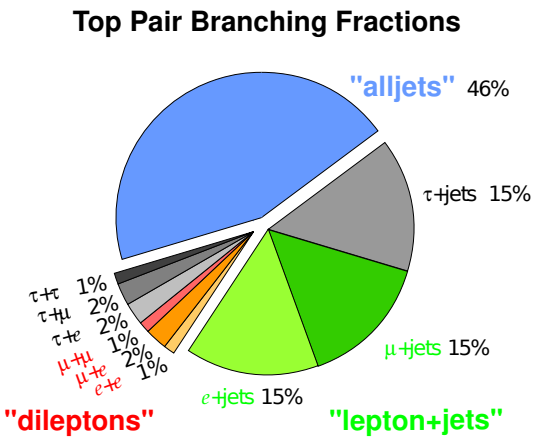


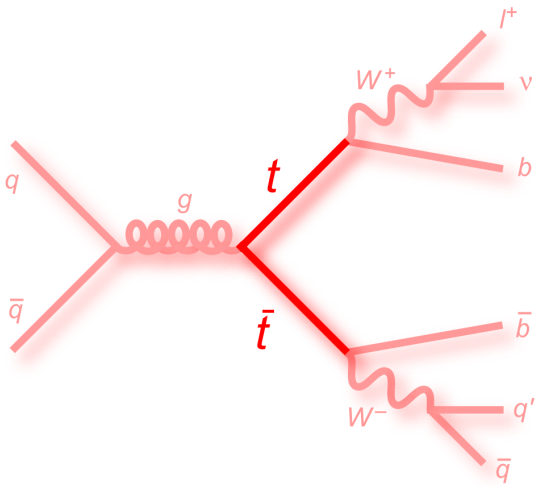
Figure 1.7: The reduced coupling strength modifiers $\kappa_F m_F/v$ for fermions (t, b, τ and μ) and $\kappa_V m_V/v$ for vector bosons (W and Z) shown as a function of their SM predicted masses and of the VEV of the Higgs field ($v = 246\,\text{GeV}$). κ_F (κ_V) describe the coupling modifiers for fermions (vector bosons) predicted by the SM to be unity. The bottom panel displays their measured values. Ultimately this displays the reduced coupling strength modifiers are directly proportional to the mass of the particles measured [6].

A third of all semi-leptonic decays involve the production of a τ lepton, which can also decay hadronically. 40% of dilepton decays produce a single tau, and 10% of dilepton decays lead to two τ leptons – in each of these cases it is also possible for the τ leptons to decay hadronically.

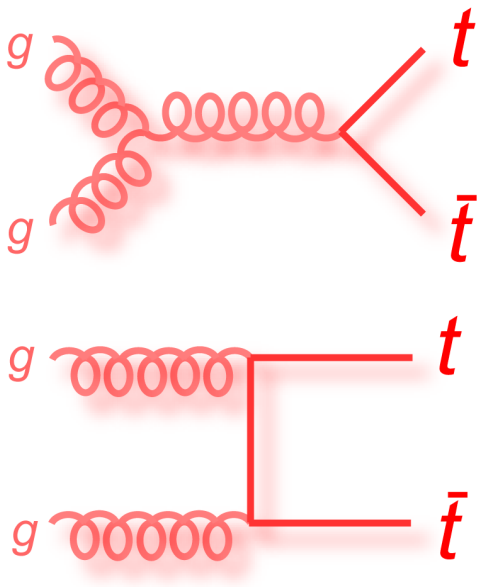
The search for $t\bar{t}H$ decays, with a cross section of approximately $670 \pm 90(stat.)^{+110}_{-100} fb$ [65], is complicated by the most common background process within the LHC, top quark pair production, $t\bar{t}$, with additional hadronic jets. With a cross section of $830 \pm 0.4(stat.) \pm 36(syst.) \pm (lumi.) pb$ $t\bar{t}$ [66] at centre-of-mass energies of $\sqrt{s} = 13\,\text{TeV}$ pair production occurs much more frequently within the LHC [67, 68, 69, 70]. 90% of these pair productions within the LHC originate from gluon-gluon fusion, with the rest initiated through quark-antiquark annihilation. Due to their



(a) $t\bar{t}$ decay branching fractions. [64]



(b) Feynman diagram of $t\bar{t}$ pair-production through quark fusion, followed by semi-leptonic decay of $t\bar{t}$ pair. [64]



(c) Feynman diagrams of two other $t\bar{t}$ pair-production processes [64]

Figure 1.8: Feynman diagrams for $t\bar{t}$ pair production and decay and diagram showing $t\bar{t}$ decay branching ratios.

multiplicity and prominence in these processes, precise facilitation and accurate understanding of hadronic jets is crucial to analysis for the Higgs boson. This is particularly true for purposes of understanding Yukawa coupling through observation of interactions between the Higgs boson and the heaviest particle in the SM, the top quark.

Higgs Decay Channel	**Branching Ratio [%]**
$H \to b\bar{b}$	58.2
$H \to WW$	21.4
$H \to gg$	8.19
$H \to \tau\tau$	6.27
$H \to c\bar{c}$	2.89
$H \to ZZ$	2.62
$H \to \gamma\gamma$	0.227
$H \to Z\gamma$	0.153
$H \to \mu\mu$	0.022

Table 1.1: Branching Ratios for SM Higgs decays [22]

THE ATLAS EXPERIMENT AT THE LHC

2

If we are to achieve things never before accomplished we must employ methods never before attempted.

Francis Bacon

This chapter will construct a technical explanation of the use of experimental apparatus for the acceleration and collision of particles. In particular, it will discuss the Large Hadron Collider (LHC), as well as the detector machinery for identifying and measuring particles, focusing on components of A Toroidal LHC ApparatuS (ATLAS).
Following this, the trigger system used by the ATLAS experiment to meet the needs of the high event rate within the LHC will be introduced.

2.1 The LHC

The Large Hadron Collider (LHC) [71, 72] is based at the European Organization for Nuclear Research (CERN), just outside Geneva, and is the biggest and most powerful particle accelerator in the world. A synchrotron style particle accelerator, with a circumference of 27 km in total, it is located in a circular tunnel beneath the Franco-Swiss border at a depth ranging between 45 m and 170 m below the surface. Within the LHC ring two beams of high energy particles travel the circumference of the accelerator in opposite directions. Most commonly these high energy particles are protons, however lead ions are also used. Each beam consists of 2808 bunches of 10^{11} protons, with each bunch only 25 ns apart. The counter-rotating beams are passed through two different beam pipes, inside an ultra-high vacuum of $\sim 10^{-10}$ mbar, and are collided at four points 40 million times a second.

The four collision points are the locations of the main experiments at the LHC, two of which, ATLAS and CMS, are complimentary, multi-purpose detectors [12, 73]. The remaining two are Large Hadron Collider beauty (LHCb) [74], which specialises in flavour physics, and A Large Ion Collider Experiment (ALICE) which specialises in heavy ion physics [75]. ATLAS and CMS are

the largest of the four, and are both designed for equivalent tasks, general-purpose searches for both the Higgs boson and signs of Beyond Standard Model physics BSM physics at the TeV scale. LHCb is constructed for precision tests of SM parameters, particularly with consideration towards CP violation. This is done through the analysis of B and D meson decays. ALICE, an asymmetric dectector, was specially built to survey the outputs of colliding lead ions with the aim of producing a quark gluon plasma and measuring properties of QCD phase transitions.

In addition to the four main detectors surrounding the ring, there are a number of smaller experiments based in the same caverns; TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC (TOTEM) [76], Monopole & Exotics Detector At the LHC (MoEDAL) [77], and Large Hadron Collider forward (LHCf) [78]. TOTEM is positioned alongside CMS and determines the total, elastic cross-section of pp collisions to compute the luminosity of the LHC [79]. These cross-sections are measured at small angles to the beam. LHCf, like TOTEM, is a forward detector and is used to examine cosmic ray shower processes by using particles scattered at small angles from the beam axis. It is situated at both sides of the ATLAS cavern. MoEDAL is used to seek evidence for stable and highly ionising massive particles or magnetic monopoles, and is located beside LHCb.

With the purpose of accelerating protons to an energy of 7 TeV, and colliding them at centre-of-mass energies of $\sqrt{s} = 14\,\text{TeV}$ the LHC was first turned on in 2008, however it has only been fully operational since November 2009 [80]. This delay was caused by an incident whereby a magnet quench damaged over 50 of the superconducting magnets contained in the complex. Following this, low energy beams were circulated for the first time. The first full data taking period, Run 1, did not start until 2010. It has since reached the highest energies for proton-proton and lead-lead collisions yet. Run 1 mostly consisted of 4 TeV beam energies, with centre-of-mass energies reaching $\sqrt{s} = 8\,\text{TeV}$. Run 1 continued until 2013 and was followed by Long Shutdown 1 (LS1), a period of no beam when required upgrades to the experiment could be undertaken. This was later than was previously planned due to the success of one of the main goals of the LHC– the discovery of the Higgs boson in 2012. In 2015 the experiment was restarted for Run 2, reaching a beam energy of 6.5 TeV, and centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Run 2 continued until late 2018 followed by the Long Shutdown 2 (LS2), which will continue until early 2021 [81].

2.1.1 Collider Physics

Particle colliders are designed to study the final state objects originating from collisions of sub-atomic particles at intersectional points known as bunch crossings. At each designated bunch crossing two conversely travelling bunches of particles (protons in the case of the LHC) will collide with an inconstant number of intersections, generating an assemblage of these final state objects. High energy physics experiments, including such detectors as ATLAS, can be used to analyse these final state objects. A collection of final state objects from a single bunch-crossing is known as a single physics event.

The LHC is a synchrotron, a category of circular collider. Circular colliders are advantageous over linear counterparts in that particle beams can be maintained in their given trajectories for extended periods of time, allowing recurrent use from multiple collisions of the same particle bunches. The major drawback to circular colliders is that a charged particle experiencing transverse acceleration, as is needed to rotate it around the ring, induces synchrotron radiation. This radiation leads to an energy loss related to the mass of the accelerated particle:

$$\frac{dE}{dt} = k\frac{E^4}{m^4R^2} \tag{2.1}$$

Where k is a dimensional constant, m is the mass of the accelerated particle, and R is the radius of curvature of the particle's trajectory. As can be seen, the energy loss is inversely proportional to m^4, and as such for light particles will be considerable. Electron and positron colliders are therefore more susceptible to the consequences of synchrotron radiation than proton colliders, where this is reduced by a factor of $(m_e/m_p)^4 \approx 10^{-12}$, and the maximum energy reach of the former is determined by this effect. For hadron colliders this is instead limited by the capability of the bending magnets and the maximum magnetic field that they can create.

2.1.2 Performance of the LHC

At the LHC, the physical processes we wish to explore result from the hard scattering of partons within the protons: quarks and gluons. The momenta of these partons is unknown, however, as they contain unidentified fractions of the total momentum of the proton [82]. There are some qualities we can determine, though.

An important value of the LHC is the luminosity of the machine [79]:

$$\mathscr{L} = f\frac{n_b N_1 N_2}{4\pi\sigma_x\sigma_y} \tag{2.2}$$

Here the values impacting the luminosity are the revolutionary frequency of the particle bunches, f, the number of particles per bunch, n_b, the number of bunches per beam $N_1 = N_2$, and the transverse area of the bunches at our interaction point, described using σ_x and σ_y, Gaussian widths representing the physical size of the beam in the horizontal and vertical direction. The luminosity is related to the total number of collisions within the machine such that

$$N_{event} = \mathscr{L}\sigma_{event} \tag{2.3}$$

Where σ_{event} is the cross-section for this event. We can also define the instantaneous luminosity at any time as

$$\mathscr{L}_{inst} \equiv \frac{1}{\sigma}\frac{dN}{dt} \tag{2.4}$$

Here σ is the total cross-section of a given physical process and dN/dt is its rate of occurrence. The measure of the total luminosity from the accelerator over a given time period,

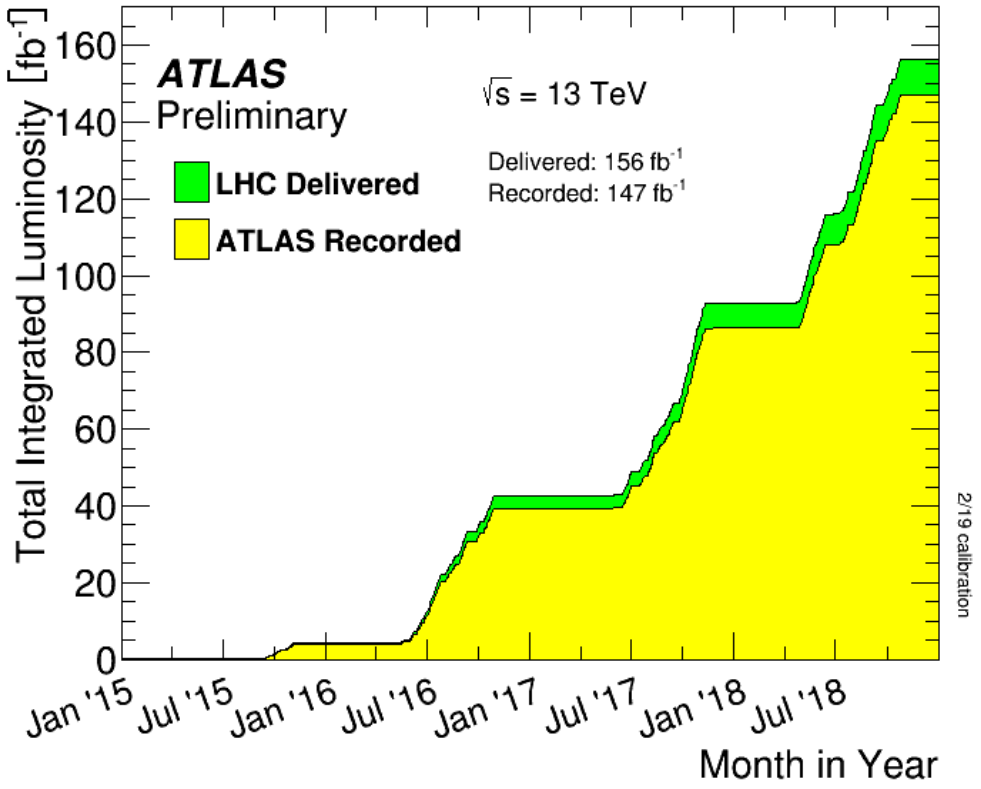


Figure 2.1: Cumulative luminosity for 13 TeV proton-proton data at the end of Run II [7]. 1. The difference between luminosity delivered by the LHC and recorded by ATLAS stems from inefficiencies in the trigger system, an inefficiencies of ramping up the tracking detectors when the LHC declares stable beams.

$\Delta t = t_2 - t_1$, is described as the integrated luminosity, defined as

$$\mathscr{L}_{tot} = \int_{t_1}^{t_2} \mathscr{L}_{inst} dt \tag{2.5}$$

We are able to determine the average number of particle interactions per bunch crossing, known as pile-up [83], $\langle\mu\rangle$ as a function of $\mathscr{L}_{inst}$:

$$\langle\mu\rangle = \frac{\sigma_{inelastic}\mathscr{L}_{inst}}{Nf} \tag{2.6}$$

Here $\sigma_{inelastic}$ is the total cross-section of inelastic scattering of the protons, N is the number of circulating bunches, and f is the bunch frequency. The average number of interactions for each bunch crossing is calculated as a mean over the specific luminosity block of interest. During Run 1 and early parts of Run 2 the rotational frequency of bunches was 20 MHz, with 50 ns between collisions. Since then it has been increased to 40 MHz, with 40 million collisions occurring each second, with 2808 bunches of up to 10^{11} protons colliding every 25 ns. The LHC was designed for beam energies up to 7 TeV, giving centre-of-mass energies of $\sqrt{s} = 14$ TeV. A peak instantaneous luminosity of $\mathscr{L}_{inst} = 2 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ was reached during early Run 2. Upon reaching the end of Run 2, the LHC had delivered 156 fb^{-1} to ATLAS. The LHC has also collided heavy ions, including lead [84].

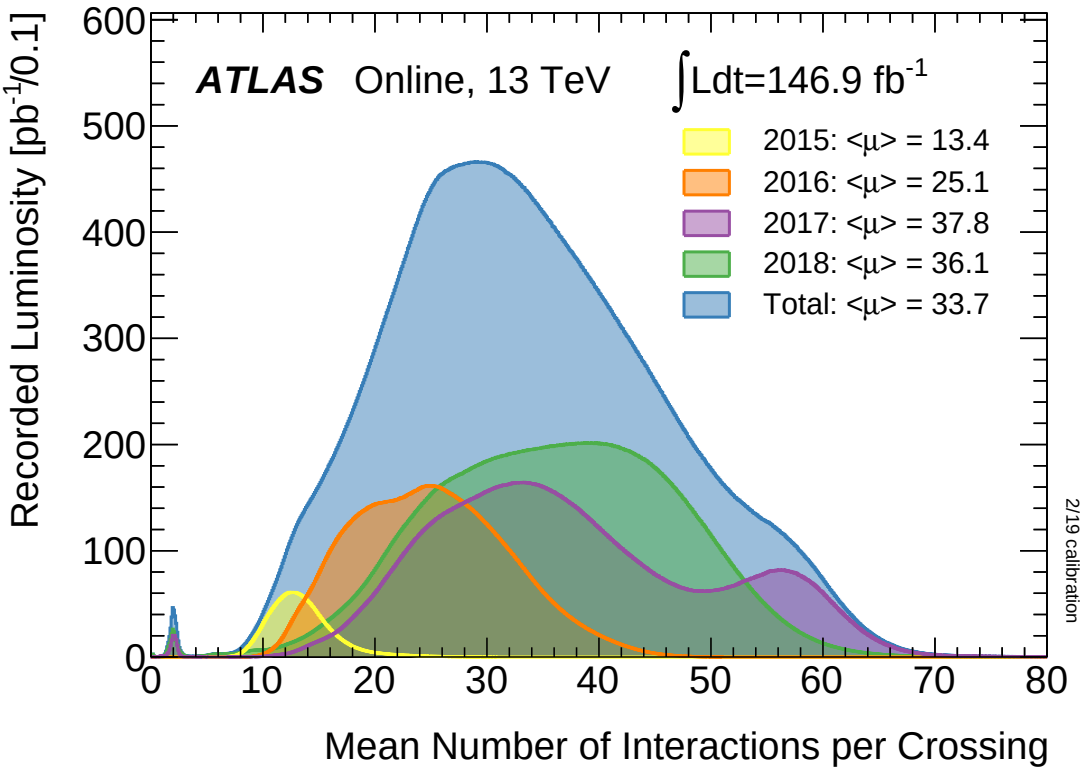


Figure 2.2: Luminosity-weighted distribution of the mean number of interactions per crossing, $\langle\mu\rangle$, for proton-proton collisions during Run 2 at CoM energy, $\sqrt{s} = 13\,\text{TeV}$ data [8].

2.1.3 Acceleration and Injection

The protons used in collisions that take place at the LHC begin their journey elsewhere in CERN, extracted from hydrogen gas [71]. Electrons are removed from the atomic hydrogen using an electric field. The remaining protons are then accelerated to higher energies and injected into the LHC through a number of intermediate steps using successively larger storage rings and accelerators (see Figure 2.3) [85]. Firstly they are accelerated by Linear Accelerator 2 (LINAC2) up to energies of 50 MeV, this is followed by the small Proton Synchrotron Booster (sPSB) to reach energies up to 1.4 GeV, then the Proton Synchrotron Booster (PSB) for energies to 28 GeV and finally the Super Proton Synchrotron (SPS) to reach energies of 450 GeV. The SPS then injects two oppositely travelling beams into the ultra-high vacuum pipes and accelerated around the LHC in opposite directions through two transfer tunnels, each about 2.5 km long.

Travelling from from LINAC2 to the LHC takes the protons about sixteen minutes, and a further twenty minutes is taken for them to reach collision energies. Within the LHC, the acceleration is accomplished by eight superconductive radiofrequency (RF) cavities: a metallic chamber with an electric field oscillating at 400 MHz. These are located at four different sites around the LHC. These RF cavities are also used to maintain bunch structure within the beam, through longitudinal beam focusing: the oscillating nature of the field in the RF sorts the beams into bunches by varying the acceleration on the proton with respect to its position within the

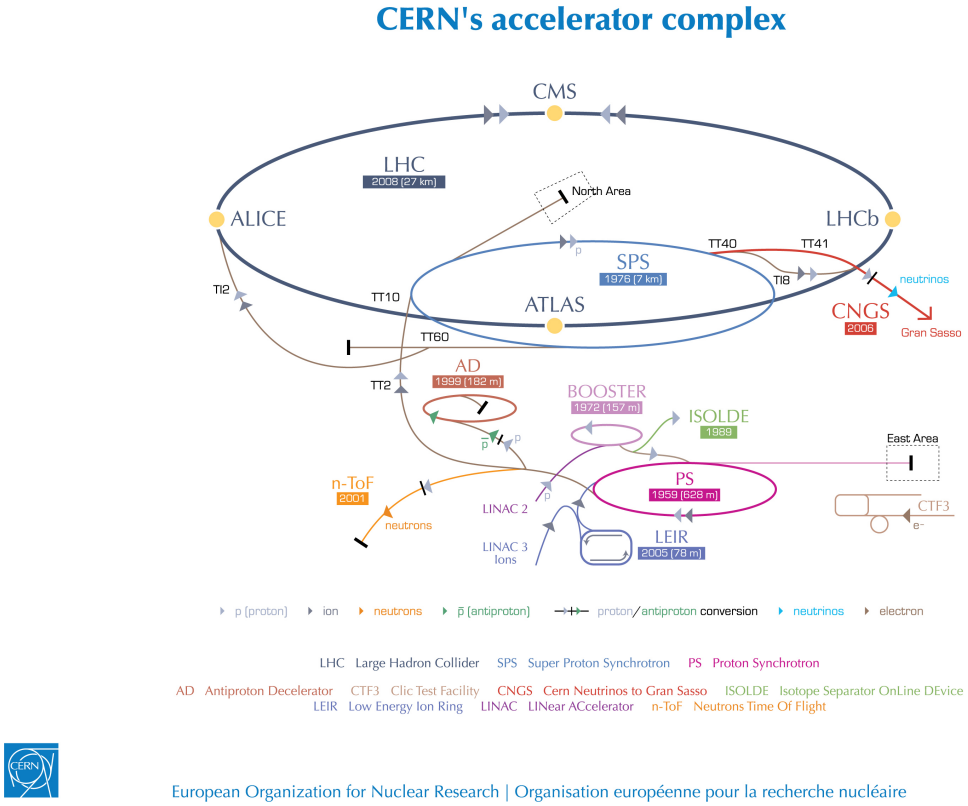


Figure 2.3: The Accelerator complex at CERN, dispaying the various accelerators and detectors [9].

bunch. Protons earlier or later than the bunch centre are decelerated or accelerated by differing degrees to keep them close in energy. Once protons reach the nominal energy the beams are able to travel around the LHC and can be stored for several hours during data taking before being needed to be replenished.

For heavy ion runs this process is somewhat different. The lead ions begin as a source of vapourised lead before being accelerated within Linear Accelerator 3 (LINAC3) and the Low Energy Ion Ring (LEIR). Following this they will pass to the Proton Synchotron and take the same path as the protons.

2.1.4 Collider Magnet Systems

Run 2 began in June 2015, after a two-year long LS1. During this shutdown the magnets were upgraded to be able to handle the requirement to circulate 13 TeV beams during Run 2. The magnet system of the LHC consists of 1232 superconducting dipole magnets, operating at a current of 11 850 A, and 392 quadrupole magnets. These magnet systems consist of special electrical cables formed into coils that can operate in a superconducting regime. They are kept supercool at a temperature of 1.7 K by superfluid helium and have an average field strength of 8.3 T.

The dipole magnets produce a transverse acceleration on the protons in the beam to maintain a circular path. This field is in the direction orthogonal to the direction of movement of

the protons. The quadrupole magnets, and a few other higher-moment magnets, are deployed to manage the beam focusing as it is accelerated around the ring, as during the beams' rotation they can diverge. The 5-7 m long quadrupole magnets refocus the width and height of the beams. Eight inner triplet magnet systems, of three quadrupole magnets each, are used to focus the beams close to the collision point. These magnet triplets lie 23 m from the interaction points. 688 smaller sextupole magnets are used to refocus the chromaticity of the beam, diverted through momentum changes in the particle bunches [86].

The momentum of the particles, p, having travelled through the magnetic field, B, in an arc of radius R can be given by the relativistic relation [87]:

$$p_T = 0.3BR \tag{2.7}$$

2.1.5 General Purpose Detectors

ATLAS and CMS, the two biggest detectors of the LHC, are both general purpose detectors - the purpose of which is to investigate the possibility of new massive particles, especially the Higgs boson [12, 73]. The need to be general purpose, and as such attempt to identify as many different final state objects as possible, is due to the unknown nature of BSM physics.

One way to improve the likelihood of success in this pursuit is to design the detectors to be fully hermetic; covering the biggest possible solid angle about the interaction region will increase the possibility of maximising observations of products of the collision. Both of these general purpose detectors consist of specialised sub-systems, designed to work complimentarily to search for different signals, and a fine granularity in order to most precisely determine the location of these signals. ATLAS and CMS rely on different technologies from one another to arrive at independent measurements of the same physical phenomena in order to cross-check and improve reliability. For this their performances are required to be comparable. If as much information can be determined from observable processes as possible, hermetic detectors can also lead to the discovery of unobserved processes through missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). This is done by simply identifying the resulting momentum imbalance in the transverse plane after reconstructing the rest of the event.

All experiments at the LHC must also meet certain general requirements. Firstly, they must all be robustly resistant to high levels of radiation, especially parts located close to the interaction region. Secondly, the detector components must be able to identify individual events at a rate of 40 MHz to keep up with the rate of bunch crossings. This leads to the need for a highly developed data acquisition process and trigger system, to select and save relevant information from events in which interesting final state objects are detected.

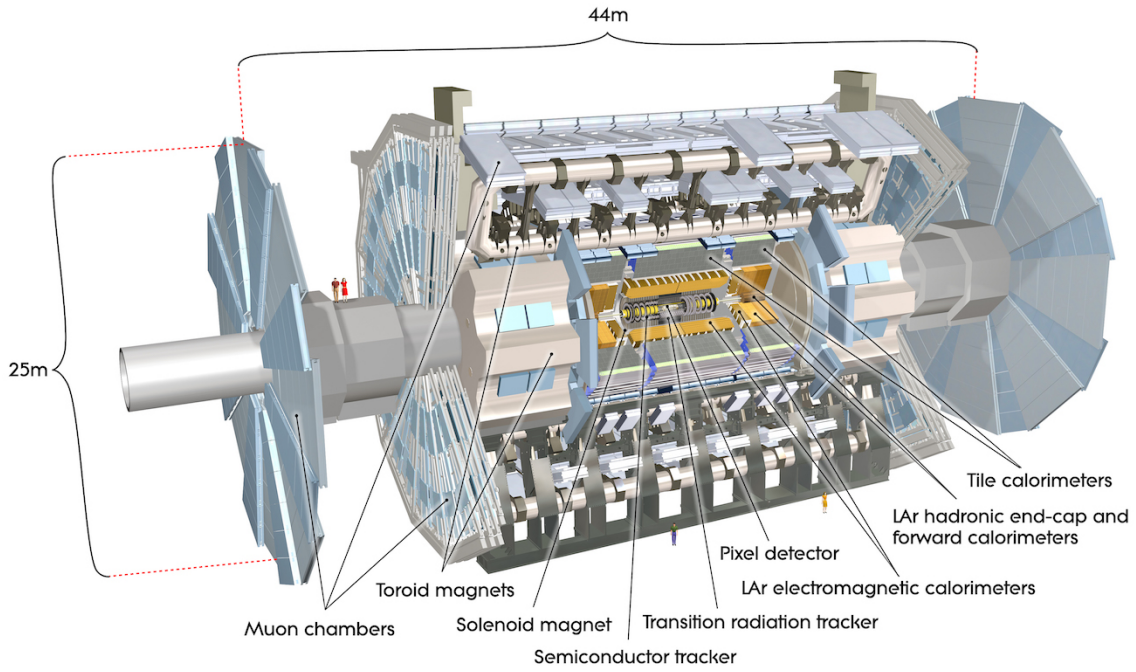


Figure 2.4: Cut-away view of the ATLAS detector and its subsystems [9].

2.2 The ATLAS detector

A Toroidal LHC ApparatuS (ATLAS) is the largest of the experiments at the LHC, and is one of the two general purpose detectors [12]. An illustration is shown in Figure 2.4. It is located approximately 100 m below ground at CERN's point 1 cavern. It has a cylindrical symmetry, and measures 45 m in length and 25 m in diameter. The full system weighs close to 7 kt.

ATLAS was designed with a wide physics program in mind, motivated by searches for the Higgs boson and new BSM physics. It has forward-backward symmetry, cylindrical geometry with respect to the interaction point, and a solid angle coverage of nearly 4π. It is designed to reconstruct and measure a range of physics objects such as hadronic jets, electrons and muons, originating from hard scattering, with special attention to objects with large transverse momentum (p_{T}) with respect to the beam axis. It has been optimised to be as sensitive as possible and to work at the highest luminosity provided by the LHC.

ATLAS consists of a series of specialised layers of subdetectors concentrically stacked surrounding the beam pipe, many of which consist themselves of modules in both the central barrel region and within the endcaps that lay at each end of the cylinder. Advancing outwards from the central beam pipe these subdetectors include a set of inner detector tracking technologies, a set of both electromagnetic and hadronic calorimeters, and a muon spectrometer as the outermost layer. Located within these layers are the components that make up the ATLAS magnets system.

The innermost section of the detector, the Inner Detector (ID), specialises in reconstructing

the trajectories of ionising charged particles and can be employed to identify production and decay vertices. It can be divided into four major components: the inner tracking detector, consisting of a silicon pixel detector; the SemiConductor Tracker (SCT) which makes use of silicon microstrips; the Transition Radiation Tracker (TRT), the third and final tracking component of the ID; and, the outermost layer, a thin superconducting solenoid magnet, which immerses the ID in a 2 T magnetic field. This magnetic field is used to curve the trajectories of charged particles within the detector to facilitate measurements of particle momentum. This will be introduced in Section 2.2.3.

The second major section of ATLAS is the pair of calorimeters, which are used to obtain precise energy measurements of electromagnetically charged particles and hadronic jets. The first of these is a fine-granularity Electromagnetic Calorimeter (ECAL), constructed of lead and Liquid Argon (LAr). The second is an iron and scintillator-tile Hadronic Calorimeter (HCAL). Both of these work in the central barrel region of the detector and are supplemented in the endcap and forward regions by LAr Forward Calorimeter (FCAL) for both electromagnetic and hadronic measurements. This will be discussed in detail in Section 2.2.4.

The final and outermost section of ATLAS is the Muon Spectrometer (MS), is used to identify and gather precision measurements of the positions and momenta of muons travelling through the detector. As well as detectors and tracking elements a set of large superconducting toroidal magnets are included within the MS to curve the trajectories of charged particles for measurements of momenta, as within the ID This will be covered in Section 2.2.5.

Each subdetector section consists of both cylindrical barrel sections, centrally placed and parallel to the beam axis, and additional discoidal “endcap” subsections at each extremity, orthogonal to the beam axis. This is to make the detector design as hermetic as possible. The barrel segments are broadly arranged with segments organised axial relative to the beam pipe, whilst the endcaps have radial divisions.

Within ATLAS indicators from the different subdetectors can be combined and used to identify the source of the signal. Photons appear as showers in the electromagnetic calorimeter, but as neutral objects leave no ionising trace in the ID. Electrons appear identical to photons within the calorimeter, but also leave tell-tale charged tracks within the ID. Neutrons, with no electromagnetic charge, leave no trace within the tracking of the ID, however will generate showers within either the hadronic or electromagnetic calorimeters. Protons add an indicatory charged track within the ID to the distinctive shower of hadrons within the calorimeter. Other hadronic, or hadronising particles can be identified through the analysis of jets detected within the hadronic calorimeters, see Chapter 4. Muons travel through most layers of the detector without much trace, but show ionising tracks within both the ID and MS. Neutrinos, distinctive in their lack of trace within the detector, can be identified by missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) after determining momentum imbalance.

2.2.1 ATLAS Geometry and Nomenclature

There exists a standardised nomenclature for the coordinate geometry used within ATLAS to give precision spatial descriptions of both kinematic measurements and detector components. ATLAS has a forward-backward symmetrical cylindrical geometry with a right-handed Cartesian coordinate system, the origin of which lies at the particle interaction point. This coordinate system starts by taking the definition of the z-axis as coincident with the direction of the beam. The side of the detector in the positive z-direction is labelled as the "A-Side", and that in the negative z-direction is the "C-Side". The x-y plane is therefore transverse to the beam-pipe, with positive x pointing from the origin towards the centre of the LHC ring and positive y pointing directly up towards the surface of the earth. The cylindrical symmetry of the detector allows us to easily introduce a cylindrical coordinate system that replaces measurements on the x-y plane with the radius, r, and the azimuthal angle, ϕ, which is measured around the beam axis. We also introduce the polar angle, θ, measured with respect to the positive-z axis of the beam pipe, to describe longitudinal positions of objects in the detector. This angle is invariant under a boost in the beam directions, however we further introduce additional coordinates labelled as rapidity, Y, and pseudorapidity, η.

Rapidity, used to describe the positions of objects with mass, such as jets, is defined as

$$Y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) \tag{2.8}$$

Where E and p_z are the energy and the z-component of momentum of the object. The sum and difference in rapidities is Lorentz invariant under boosts in the z-direction. This is useful at ATLAS as the unknown and variable fraction of momentum carried by partons within the proton lead to unknown longitudinal boosts. Particles at the LHC are assumed to be highly relativistic with negligible masses, however, so a new spatial coordinate is introduced: Pseudorapidity, η, which is also used to describe the angle of a particle with respect to the z-axis. Defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.9}$$

with $|\eta| = 0$ corresponding to the centre of the detector (i. e.where $\theta = \frac{\pi}{2}$) and $|\eta| = \pm\infty$ to the forward regions (where $\theta = 0;\pi$). The distance between two points in psuedorapidity-azimuthal angle space is then defined as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{2.10}$$

ΔR is often used in the situation of imposing a check for spacial proximity between two event objects, whereby the requirement

$$\Delta R < R_{max} \tag{2.11}$$

defines a space of radius R_{max} about the first object within which the second object must reside. This is often used in the case of "matching" objects to one another.

An important kinematic variable that is commonly used in ATLAS is the transverse momentum (p_{T}) of an object, which is measured in the x-y plane and is defined as

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{2.12}$$

Here p_x and p_y refer to the components of the object's momentum in the x- and y-direction. p_{T} is particularly valuable, as the magnitude of the longitudinal momentum, $|p_z|$, of a product of our collision depends on the unknown momenta of the partons within our colliding protons. The initial p_{T} is known to be zero, however, so this can be used to determine missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). If the final visible p_{T} of the system, p_{T}^{vis}, is non-zero, then we can determine, due to momentum conservation, the presence of $E_{\mathrm{T}}^{\mathrm{miss}}$, caused by some unknown component.

$$E_{\mathrm{T}}^{\mathrm{miss}} = \left|\overrightarrow{p_{\mathrm{T}}}^{miss}\right| = \left|\overrightarrow{p_{\mathrm{T}}}^{vis}\right| = \left|\sum_{i=1}^{i} \overrightarrow{p_{\mathrm{T}}}^{vis,i}\right| \tag{2.13}$$

Where $\sum_{i=1}^{i} \overrightarrow{p_{\mathrm{T}}}^{vis,i}$ is the measured summation of momenta from all visible objects in the event.

2.2.2 Magnet System

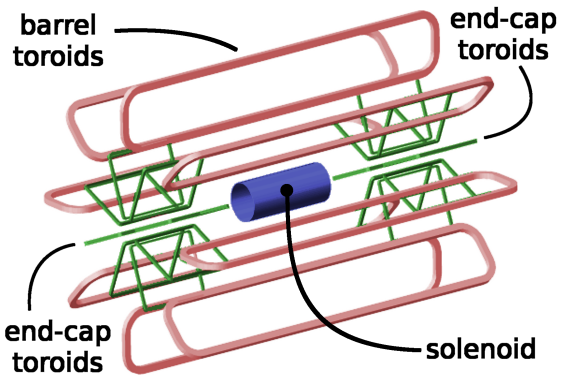


Figure 2.5: Schematic of the magnet system within the ATLAS detector [10].

One of the essential features of the ATLAS detector is the magnet system, composed of four large superconducting magnets [10]. At 22 m in diameter and 2.6 m in length the magnets use a stored energy of 1.6 GJ to deflect the trajectories of charged particles in order to perform momenta measurements. The direction and radius of curvature of the charged particles' path subsequent to deflection by the field is directly related to their momenta and charges.

The magnet system consists of three subsystems: a single solenoid, one barrel toroid, and two toroidal endcaps. Each of these is formed of NbTi, a superconducting material with a critical temperature of 10 K, and cooled by liquid helium to 4.5 K by the magnet cryostat system. It is designed to provide a precise and stable magnetic field.

Central Solenoid

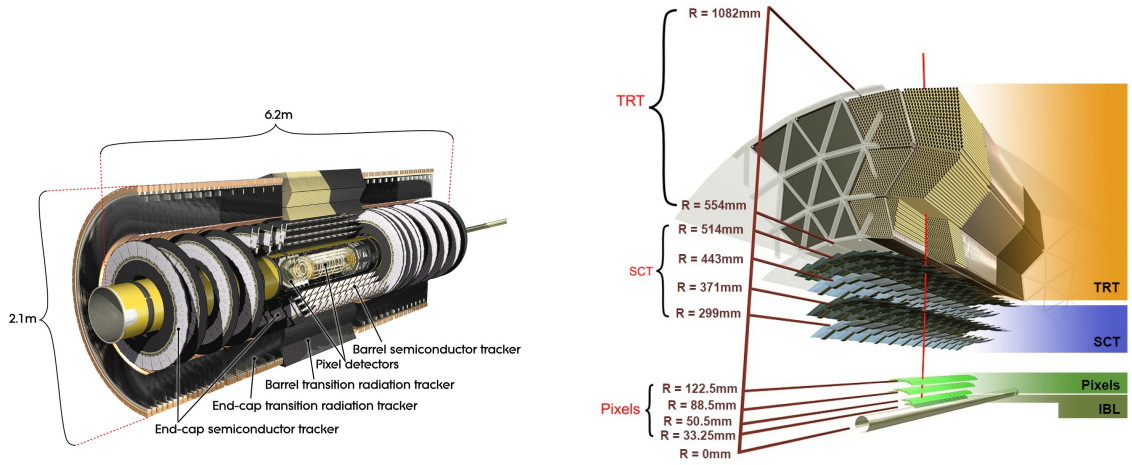
The Central Solenoid is located between the ID and the barrel ECAL and is designed to provide a 2 T axial magnetic field with a minimal radiative thickness in front of the calorimeters [88].

Despite its thinness - with an inner radius of 2.46 m and outer radius of 2.56 m - it still has an axial length of 5.8 m and can allow accurate momentum measurements up to 100 GeV.

The Barrel and the Endcap Toroids

The two toroidal subsections of the magnet system - the air-core barrel toroid and the pair of air-cooled endcap toroids - provide a magnetic field of 3-8 T for the MS. The centrally-placed barrel toroid is comprised of eight coils and has an axial length of 25.3 m, an inner diameter of 9.4 m and an outer diameter of 20.1 m. The forward-placed endcaps also comprise of eight coils each, are 5 m in length and have inner and outer diameters of 1.65 m and 10.7 m. The magnetic field produced here is orthogonal to the particle direction within the MS and is used to measurement muon momenta.

2.2.3 Inner Detector



(a) Cut-away schematic of the ATLAS ID. Note the IBL is not shown.

(b) A crosssection of the layered sub-detectors of the ATLAS ID.

Figure 2.6: The ATLAS ID [11].

The Inner Detector (ID) is the closest sub-detector to the central beam pipe, and is therefore first to measure products of the collision, including decays of short lived particles [89]. The ID measures 6.2 m in length and 2.1 m in diameter and is enveloped in a 2 T magnetic field produced by the solenoid magnet. The ID is designed to provide reliable identifications and measurements of momenta, and both primary and secondary vertices.

There are a number of restrictions placed on the technology of the ID. The depth of material needs to be thin in order to limit the interactions of particles before they reach the calorimeters, which would lead to a degradation of quality in the energy measurements. The close proximity of the sub-system in relation to the beam means that there are extremely high levels of radiation that the ID must contend with; the equipment within must be resistant to damage from this. Finally the vast multiplicity of particles formed within proton collisions means that all the

parts of the ID need to produce measurements with a high level of precision, so as to determine individual tracks from a series of positional "hits" – this means that the subsections of the ID must be finely segmented to increase granularity.

To overcome these complications, the ID is composed of multiple specialised sections arranged into three separate elements: a main barrel with concentric cylindrical layers, and two disk endcaps, to ensure that it covers all particles in a range of $|\eta| < 2.5$. There are four main sections of the ID: an Insertable B-Layer (IBL) and Silicon Pixel Detector (SPD) used to accurately measure and reconstruct primary and secondary vertices through silicon pixel layers; the SemiConductor Tracker (SCT), a layer used for accurate measurements of particle momenta, constructed from silicon microstrip layers; and, finally, the Transition Radiation Tracker (TRT), which provides continuous tracking and electron identification complementarily to that of the ECAL over a wide range of energies. A detailed illustration can be see in Figure 2.6

High-resolution spacial measurements supplied by these sections of the ID are used in algorithms to reconstruct the paths of particles, called tracks.

Insertable B-layer

The innermost section of the ID is the Insertable B-Layer (IBL) [90] [91]. Added during the Run 2 upgrade, the addition of this new layer was done to improve Primary Vertex reconstruction by a factor of 1.4 (important for the tagging of bottom-quark-initiated jets), double the quality of impact parameter tracking precision by supplying an additional space-point, and, also, to prevent damage to the inner layers due to higher radiation levels. It is comprised of 6 million channels, has a spatial resolution of 8 µm by 40 µm and is the closest section to the beam axis, at only 33.25 mm distance.

Silicon Pixel Detector

The second layer of the ID is the Silicon Pixel Detector (SPD), the main feature of which is its fine granularity, with a resolution of 10 µm by 115 µm, necessary for precision primary- and secondary-vertex reconstruction and measurements [92]. The SPD is made of three concentric layers of silicon sensors that surround the beam axis, and a total of six disk layers, with three at each forward region. The barrel section is cylinder of length 6.2 m and diameter of 48.4 cm that extends out to a pseudorapidity of $|\eta| < 2.5$. The endcap disks are all mounted perpendicular to the beam axis to track charged particles at high η.

Both the SPD's barrel and endcaps consist of three concentric layers themselves: the b-layer, or L0, at 50.5 mm; L1 at 88.5 mm; and L2 at 122.5 mm. All of the barrel's 1,456 modules, and 288 in each of the endcaps, has 46,080 readout channels, or pixels. The smallest of these pixels has a surface area of 50 µm in the ϕ direction by 400 µm along the z-axis. Totalling about 80 million rectangular segments of silicon sensors, or pixels

Each of these three layers will be traversed by ionising charged particles, leaving "hits" as

they do so. The ionisation of the semiconductor material produces electron-hole pairs, which are channelled by an electric field so the resulting current can be detected. This allows the identification of "space-points" traversed by the particle. Each pixel also has an independent electronics channel with a separate circuit.

Semiconductor Tracker

The third subsection of the ID is the SemiConductor Tracker (SCT), designed to give precision momentum reconstruction through four precise spacial measurements of charged particle tracks [93, 94, 95, 96]. The SCT is arranged in one barrel region, that extends to $|\eta| < 1.7$ and two endcaps that reach from $1.2 < |\eta| < 2.5$. Each of these sections is constructed of silicon microstrip layers which record two “hits”, one on each side of the module, as it is crossed by a charged particle. The pair of hits is combined to make a single space point allowing the measurement of track momenta, vertex positions and impact parameters.

There are 4088 modules in total. The barrel region includes four concentric layers of these silicon detectors between a radius of 299 mm to 514 mm with a total of 2,112 modules. The two endcap sections have nine layers each and a total of 1,976 modules. Each of these modules comprises of two sides of 786 back-to-back silicon strip detectors at a stereo angle of 40 mrad to one another and set at an average pitch of 80 µm to provide precision measurements of a particle's position in ϕ.

The components of the SCT are made of strips of silicon, 80 µm by 12 cm, to cover a larger area of space when compared to the two inner layers of the ID. Within the barrel region these strips are semi-parallel to the axis of the beam direction, whilst in the endcaps they are arranged radially. The geometry of the SCT is designed to achieve a precision measurement of the position on hits in the longitudinal direction, even with the length of the strips. A resolution of 17 µm in the radial plane, by 580 µm in the direction z-direction is obtained.

The silicon strip sensors are read out by a radiation-hard front-end chip, with each chip reading out 128 channels. As the SCT is further away from the beam pipe compared to inner layers there is a reduced particle density expected, this allows lower levels of granularity to be used within whilst maintaining a high level of performance. There are about 6.3 million readout channels (about 2 million fewer than the pixel detector).

Transition Radiation Tracker

The final layer of the ID is the Transition Radiation Tracker (TRT), which aims to improve p_{T} resolution for particles with longer track lengths [97, 98]. It is cylindrical and comprised of layers of nearly 300,000 plastic gaseous straw tube elements surrounded by transition radiation material. It consists of a barrel section and two endcap sections and extends to $|\eta| < 2.0$.

The TRT barrel segment is made of three concentric layers, each with 32 modules, with radii between 55.4 cm and 108.2 cm. Each layer contains approximately 50,000 straws, each 1.44 m

long and aligned with the beam axis with independent readouts at each end. It extends over a region of $|\eta| < 1.0$. Both of the endcap regions are divided into 18 wheels, with 224 layers, and a total of around 320,000 radially arranged straws. These cover a region of $0.8 < |\eta| < 2.0$. This overlap means that there can be hits in both the barrel and the endcap.

The tube elements are straws of plastic, wound from a multilayer film reinforced with carbon fibres, and are 4 mm in diameter. Each straw is filled with a gas containing 70% Xe, for x-ray absorption, 27% CO_2, and 3% O_2, both to increase the electron drift velocity and photon quenching. A 30 µm gold-plated tungsten wire runs through the middle of each straw to be used as an anode, and the inside face of the straw is coasted with aluminium, a high-voltage cathode. Each wire is divided into two halves, close to $\eta = 0$. The gas contained within the straws operates as a drift chamber would. It is ionised as charged particles traverse it and electrons and ions are collected at the anode or cathode. This means that a current is created within the straw and, due to the nature of the electric field within the tube being known, the time of electron drift to the wire can be used to determine the distance between the wire and the path of the charged particle.

Additionally between the straw elements is a transition radiation material. This consists of polypropylene fibres in the barrel region and polypropylene foils within the endcaps. Transition radiation occurs as a charged particle passes between media with different dielectric constants.

Within the central region of the TRT there are typically about 35 hits per subsystem for a single charged particle passing through, significantly higher than the SPD and SCT, with 3 and 4 hits respectively. It has a spacial resolution of 130 µm. This allows for improved estimation of track parameters when combined with the SPD and SCT.

The TRT allows the identification of particles through the detection of x-ray photons which are emitted as highly relativistically charged particle cross the media boundaries [99]. The Radiation at x-ray frequencies is recorded as an additional, high-threshold hit and is proportional to the Lorentz factor of the charged particle, defined as

$$\gamma = \frac{E}{mc^2} \tag{2.14}$$

Electrons and charged pions can be distinguished for a p_T range of 1-150 GeV using the ratio of high threshold hits, as γ is significantly larger for electrons than it is for heavier particles. The low threshold required to identify a single hit is 300 eV, whereas the high-threshold is between 6 and 7 keV. There is a significant chance of high threshold hits for electrons of approximately 1 GeV, and upwards of 200 GeV for other, heavier, particles. Measurements of (dE/dx) can also be used in identifying protons and kaons with E_T^{miss} less than 10 GeV.

2.2.4 Calorimeters

The second major sub-detector elements of ATLAS are the calorimetry sections, shown in Figure 2.7. Located outside the 2 T solenoid magnet, it consists of detectors with full symmetry in

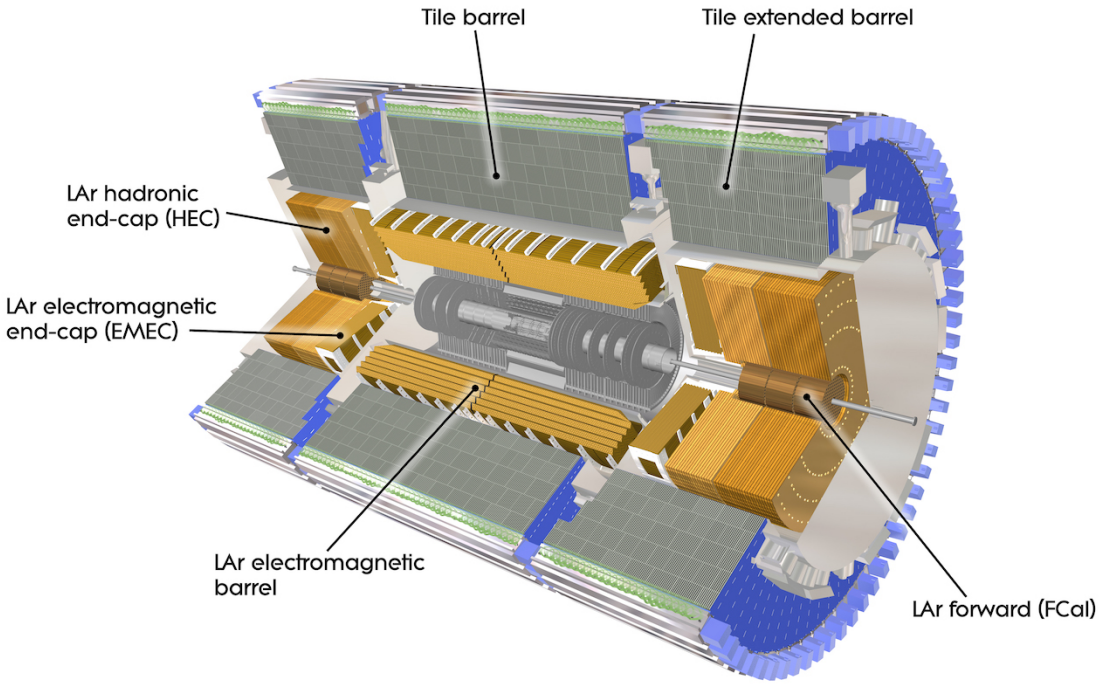


Figure 2.7: The ATLAS Calorimeter, including the ECAL, HCAL and FCAL sub-components [12].

ϕ, and a coverage of $|\eta| < 4.95$. The calorimeters have been designed to fully absorb particles and to translate their energies into measurable variables, and also to be hermetic, allowing a measurement of the total event as accurately as possible.

There are three main sub-sections of the Calorimeters: the Electromagnetic Calorimeter (ECAL), which detects and measures electromagnetically interacting electrons and photons with a fine granularity; The Hadronic Calorimeter (HCAL), which has a courser granularity and is used to identify and measure particles that interact via the strong force; and The Forward Calorimeter (FCAL), which absorbs both types of particles that have trajectories close to the beam pipe. The ECAL is the innermost section and measures electron and photon energy loss through a cascade of electromagnetic shower interactions through the calorimeter. It consists of a Liquid Argon (LAr) barrel, and two endcap sections. The LAr is used as a sensing element, where the showers taking place in the Argon will free electrons that are then collected and measured. These electromagnetic showers occur each time a high-energy photon or electron passes through the material. Below a few MeV the dominant effect of a photon passing through a material is Compton Scattering, or the photoelectric effect. At energies above this the photon primarily interacts through pair-production, creating electron pairs. High-energy electrons, including those created in the pair-production, emit photons via bremsstrahlung. This cycle of photon and electron-pair-creation, the shower, continues until the energy of the emitted photons are below the pair-production threshold, whereby energy loss of electrons starts to dominate.

The HCAL comprises of two major sections: one central section containing the Tile Barrel

and Tile Extended Barrel, known as the "Tile Calorimeters", and a section in the forward region consisting of the two endcaps, which use LAr as a sensing element. The sensing material used for the barrel sections is a collection of tiles made of scintillating plastic. The light produced by these tiles when hadronic particles interact is detected and measured. The FCAL also uses LAr in the forward region, and aims to cover the section closes to the beam axis.

Each of the calorimeters uses different technologies to target both electromagnetic and hadronic particles. They are non-compensating, meaning their signal responses to the electromagnetic and hadronic energy release are different, therefore a dedicated correction must be applied to calibrate the different types of showers at the same energy scale. The only particles known to not interact with the calorimeters are neutrinos.

All of the calorimeters rely on sampling technology, in which the material that absorbs the incoming particles is distinct and different from the material that is used to measure their energy [100]. The first layer of the calorimeters, the "absorber", is a high-density absorbing material used to initiate energy loss by producing a shower of lower-energy secondary particles. These lower-energy secondary particles then continue onwards to the second active material layer, the "sampler", which measures the progressively degrading energy, either through use of scintillation or ionisation. The calorimeters are also designed to ensure they can contain the developing showers, improving energy measurements and also preventing particle showers "punching through" into the muon spectrometer.

An important factor involved in the design of the calorimeters is the radiation length of the material used, X_o, which corresponds to the distance taken to reduce the energy of a single electron by a factor of $1/e$. This length is used to define the thickness of the ECAL. The equivalent for the HCAL is the nuclear interaction length, λ_i, the mean distance travelled by a hadron before being subjected to an inelastic interaction. On average the nuclear length is on the order of one magnitude larger than the average radiation length, hence hadronic particles are much more penetrating within ATLAS than electromagnetic particles.

The Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is the innermost section of the calorimeters and lies outside the solenoid magnet [101, 102]. In order to reduce the amount of dead material in front of the calorimeters, they share a common vacuum vessel. The aim of the ECAL is to accurately and efficiently identify photons and electrons over a wide range of energies, from about 5 GeV to 5000 GeV, and to measure their energies.

The calorimeter is divided into a central barrel, covering a region of $|\eta| < 1.475$ and two endcaps with a range of $1.375 < |\eta| < 3.2$. The barrel is divided into two half-barrel wheels and is housed in the barrel cryostat. The endcaps are located in endcap cryostats. Additional to this, there is a region defined as the crack-region, found at the meeting space of the barrel and endcaps. Signal from this region is discarded in analysis due to the large volume of material obscuring the detectors.

The ECAL is a sampling calorimeter, using lead (Pb) plates as the absorbing material and Liquid Argon (LAr) ionisation chambers with copper electrodes as the active sampling component. Within the barrel these modules are arranged radially and placed in alternating layers in a folded accordion geometry. This means they provide a full azimuthal (ϕ) coverage, maintain an even energy resolution throughout , give fast extraction of the signal from the front or rear end electrodes and also help to avoid the presence of radial cracks of non-interactive material. The decision to use LAr is based on its linear response combined with large yields from signals, as well as its being robust against radiation damage. In the endcaps the layers are parallel to the radial direction and run axially.

The functionality of the calorimeters relies on the LAr becoming ionised by electromagnetic showers traversing the material. Bremsstrahlung, Compton Scattering and electron pair-production from photons all occur from interactions taking place in the material and give rise to electromagnetic showers. These showers ionise the LAr in active regions, and the resulting ionisation charges can be detected.

The LAr within the ECAL is cooled to 90 K. Liberated electrons within the material are collected and recorded by 101,760 copper electrode readout channels in the barrel, and 62,208 in each of the endcaps. The read out circuits are made of three copper layers insulated by two layers of polyimide. The two outermost layers of the circuitry are split into sectors, connected to high-voltage sources, polarising the LAr gap to the absorber. The inner layer is where the signal is collected through capacitive-coupling, which is then segmented into read-out pads.

The ECAL starts with a presampling later, a thin layer of active LAr, inside the barrel cryostat, with 10,880 readout channels. This presampler is able to produce a measurement of the energy lost in the dead material of the support structure upstream of the ECAL, such as the supporting walls of the cryostat. This allows the correction for energy lost by taking a measurement just before the majority of the electromagnetic showers is developed. This layer is only 0.5 cm thick in the endcaps and 1.1 cm in the barrel region, and covers a range of $|\eta| < 1.8$. The granularity of the barrel presampler is very small, at just $\eta \approx 0.003$, allowing precision pointing of photons.

The first sampling layer of the Calorimeter is designed to bring precision measurements, including distinguishing between single-shower prompt photons, and those from neutral pion decays with double showers. It is necessary for it to have the smallest measurable segment size of the three main layers at $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.

The second sampling layer contains the majority of the electromagnetic shower, the largest fraction of energy deposited, and is used for the main energy measurements. High-energy showers can extend to the third sampling layer. Since hadronic showers generally deposit more energy and are more penetrating the absence of deposits this layer is used to identify EM showers. As a layer with more focus on discrimination it has a coarser granularity at $\Delta\eta \times \Delta\phi = 0.05 \times 0.025$.

The depth of the three sampling layers is highly optimised, motivated in large part by rejection of neutral pions, π^0. The first layer has a depth, including dead material and presampler, of $6X_o$. The second layer extends to a depth of $22X_o$ and the third layer extends between a depth

of $2X_o$ and $12X_o$.

The Hadronic Calorimeter

The next part of the calorimetry is the Hadronic Calorimeter (HCAL), surrounding the ECAL [103, 104]. The main purpose of the HCAL is to provide energy measurements of hadronic objects within ATLAS. It is constructed of steel and scintillating tiles coupled to optical fibres, which are read out by photo-multipliers. The tiles are 3 mm thick, positioned perpendicular to the beam axis, and are staggered in depth.

The HCAL barrel section is segmented into three cylindrical layers with an inner radius of 2.28 m and an outer radius of 4.23 m. The three sections are the central barrel, which is 5.65 m long and covers a region of $|\eta| < 1.0$, and the two moveable extended barrels, both 2.9 m long and covering a range of $0.8 < |\eta| < 1.7$. There are also two endcaps located directly behind the ECAL endcaps, sharing the same LAr cryostat, and covering a region of $1.5 < |\eta| < 3.2$. The endcaps are both 1.78 m in length along the beam pipe, and 2.03 m in radius. The endcaps are positioned to overlap with the extended barrels to account for the drop in material density in these locations.

Each of the cylinder sections consists of 64 modules spread in the ϕ direction. Each module is further divided radially, into three layers with a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for the two innermost layers and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ for the outermost. The three layers are 1.5, 4.1 and $1.8\lambda_i$ thick in the central barrel and 1.5, 2.6 and $3.3\lambda_i$ thick in the extended barrels. Module layers are offset with respect to the layers of their neighbours to increase granularity. Each hadronic endcap consists of two wheels, each containing two layers of wedge-shaped modules which use copper absorber layers alternated with LAr.

The materials and techniques used in the HCAL are different from those used within the ECAL and vary within the tile calorimeter itself. Steel is used as the absorbing material within the barrels, and is interleaved with plastic scintillating tiles as the active medium. The scintillating tiles are read out from both sides by wavelength-shifting fibres to photomultiplier tubes. 5,760 of these are located in the main barrel and 4,092 in each of the extended barrels. The endcaps use copper plates as their absorbing material and LAr as their active material and are connected to 5,632 readout channels each. The electronics of the HCAL are contained within supportive plastic girders. These girders also provide flux return for the central solenoid magnetic field.

The Forward Calorimeter

The final part of the calorimetry is the Forward Calorimeter (FCAL) [105]. The LAr FCAL covers a range of $3.1 < |\eta| < 4.9$ and is placed at a distance of 4.7 m from the collision point. It is formed by three modules in total. Firstly, one electromagnetic layer with copper as a shower-initiating metal. Following this are two hadronic layers using tungsten as absorbers. These are also employed to provide containment and minimise lateral spread of hadronic showers. The absorber metal matrix of regularly spaced longitudinal channels is filled with concentric

rod and tube electrodes and small gaps are filled by LAr. To avoid problems with ion build up caused by higher particle fluxes, the LAr gaps in the FCAL are made substantially smaller than in other parts of the calorimeter (0.25 mm compared to 8.5 mm in the endcaps). The rods are at a high, positive voltage, and the tubes and matrix are both grounded. Each module of the three modules is 45 cm long.

Measurements within the FCAL are not as accurate as those from the other components of the calorimeter system, due to a coarser segmentation and to the large amount of background from underlying hadronic activity in the forward region. There are 1,008 readout channels within the electromagnetic module, and 754 channels in the hadronic modules.

2.2.5 Muon Spectrometer

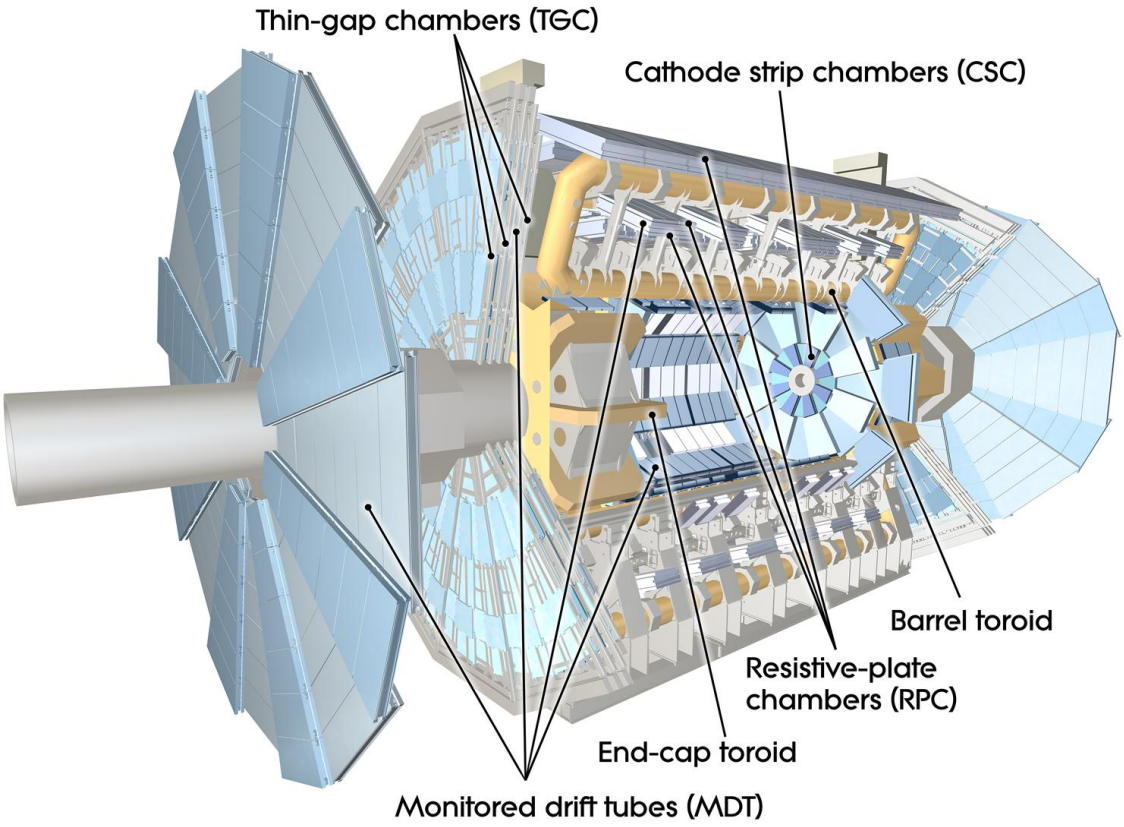


Figure 2.8: Overview of the ATLAS MS [12].

The outermost detecting section of the ATLAS detector, the Muon Spectrometer (MS), shown in Figure 2.8, surrounds the calorimeters and is used to measure the most penetrating particles detected by ATLAS: Muons [106]. Muons lose far less energy through interaction processes, such as Bremsstrahlung, than electrons due to their higher mass. They do not interact strongly and their electromagnetic energy loss through Bremsstrahlung is suppressed relative to that of an electron by a factor of $(m_e/m_\mu)^4 \approx 5.4e^{-10}$. Their lifetime within the detector is also measured as longer, due to time dilation processes originating from their relativistic momenta at the LHC.

Muons do leave a track in the ID, small energy deposits in the Calorimeters, and then go onto leave additional track segments in the MS.

The MS is used to determine the paths of muons taken through the detector, and also to deliver precision measurements of their momenta. The design of the MS is similar to the inner detector, it employs large superconducting air-core toroid magnets to cause deflection of the muons. The muon trajectory in the MS is curved in the $R-z$ plane. Measurements of their trajectory within a magnetic field can be taken through the use of high precision tracking chambers to accurately measure their momenta. Measurements can also be made of their momentum, direction and electric charge. The larger size of the MS allows accurate measurements of muons up to the TeV scale, which is essential for the physics programme of ATLAS. The MS can also act as a stand-alone trigger for muons with p_{T} of the order of several GeV.

Distinct from the other ATLAS sub-detectors, the MS has its own separate components for fast muon-triggering online, and precision muon-reconstruction offline. It has four different types of component channel: two for precision measurement and two for triggering, as the readout time of the precision measurement chambers is longer than the required limit for triggering. (After a muon has been detected with the precision chambers the other detector subcomponents would no longer be storing the data from the corresponding bunch-crossing due to the amount of time the precision chambers require to record data.) These trigger chambers are operational in the region of $|\eta| < 2.4$, with two different types of chamber employed to handle the increased rates in the forward regions. The muon trigger system selects interesting events containing muon candidates by providing identification of the individual bunch-crossings and measuring the muon track in the ϕ plane which is orthogonal to the one measured by the tracking chambers. The muon trigger system defines Region of Interest (RoI) in η and ϕ which are then scanned with precision by Monitored Drift Tubes (MDT).

The magnet system used to bend the muon tracks consists of two sections; a single large barrel toroid covering the region $|\eta| < 1.4$, and two endcap magnets at either end of the barrel, covering the regions $1.6 < |\eta| < 2.7$. The region between these, $1.4 < |\eta| < 1.6$, is the transition region and is covered by both the barrel and endcap magnets. In the barrel region the toroidal field is produced by eight very large, superconducting coils and the muon tracks are measured using chambers arranged into three cylindrical layers about the z-axis. The layers lie at a radius of 5, 7.5 and 10 m. Within the endcap regions and the transition region, the chambers are installed in the R-ϕ plane, also in three layers. There are four endcap wheels at a longitudinal distance of 7.4 m, 10.8 m, 14 m and 21.5 m from the centre of the detector.

Including the MDT there are two drift chambers. The other is the Cathode Strip Chambers (CSC). These are both dedicated to high-resolution measurements of tracks in the pseudorapidity range $|\eta| < 2.7$. Over most of the η-range, a precision measurement of the track coordinates in the principle bending direction of the magnetic field is provided by the MDT. CSC, which have a higher granularity than the MDT, are used for the innermost plane over $2 < |\eta| < 2.7$. The drift chambers contain gas which is ionised by the passing muon, and ionisation electrons are then attracted to, and read out by either one of the MDT or several CSC

wires. The time taken for electrons to reach the wire is known as the drift time. The drift times in the MDT and CSC chambers (approximately 100 ns), are too long to allow the use of MDT and CSC in the first trigger level, which requires the detector to be read out in less than 25 ns (the bunch crossing time).

The dedicated trigger subdetectors consist of the Resistive-Plate Chambers (RPC) in the barrel, $|\eta| < 1.05$ and Thin-Gap Chambers (TGC) in the higher background region of the endcaps, $1.05 < |\eta| < 2.7$. These are used to complement the precision chambers and to attach measured signals to certain bunch crossings. They provide a more coarse measurement of η and ϕ , but have a response time of less than 25 ns, making them appropriate for use in hardware trigger decisions. The RPC use parallel electrode-plates rather than wires to detect muons, while the TGC are multi-wire proportional chambers. In addition to their triggering capability, the RPC and TGC are used to measure the curvature from the toroidal magnetic field, which is in the $R-z$ plane. These also determine a second reading of the muon coordinate which is orthogonal and complementary to that taken by the precision chambers. The RPC have a spatial resolution of 1 cm and a very fast response time of about 1 ns which is necessary for the triggering. The resolution of the TGC is slightly better at 5 mm and they have a response time of 4 ns.

There are several sections within the MS that have fewer chambers. Close to $\eta \approx 0$ there are control systems for the ID, calorimeters and solenoid magnet. At the base of the MS there are 9 rows of "feet" used to support the entire ATLAS detector. In all other regions there is a pairing between each RPC and MDT layer. The lack of space near the feet requires smaller RPC without MDT. In order to achieve the required performance it is important that the chambers are aligned accurately and that their position is well known. For this reason the muon system contains an optical alignment system both in and between the muon chambers, and this is complemented by track-based alignment.

Monitored Drift Tubes

The first section of the MS are is the Monitored Drift Tubes (MDT). A precision measurement drift chamber, the MDT are aluminium tubes filled with gas, and provide momentum measurements in the $|\eta| < 2.7$ region of ATLAS. In the innermost endcap region they only reach $|\eta| < 2.0$. The tubes are 29.970 mm in diameter, and are filled with Argon (93%) and CO_2 (7%) gas. Through the centre runs a tungsten-rhenium wire held at a constant potential of 3080 V.

When a penetrating muon reaches these tubes it results in a trail of electrically charged ions and electrons, which can then drift to the sides and the centre of the tube to be collected for measurements by the anode wire. By measuring the drift time of these charged electrons and ions from their initial point, the position of the muon can be determined. The major shortcoming of the MDT is that their drift time can reach up to 700 ns, limiting the maximum rate of operation of the system. Tubes have an average precision of 80 µm, enhanced by having multiple layers of tubes for each module. Chambers have a resolution of 35 µm.

Within the barrel, the MDT are arranged axially, and in the endcaps they are arranged radially, moving outwards from the beam axis. They are disposed orthogonally with respect to the beam axis, and they only provide a measurement of the coordinate of the hits. They able to take a total of twenty measurements for each track in both the barrel and endcaps.

Cathode Strip Chambers

The second part of the MS high precision measurement region are the Cathode Strip Chambers (CSC). The use of faster detectors is needed to cope with the demanding rates and background conditions in the forward region. These work in a similar fashion to the MDT, but instead of tubes with central wires, there are cathode strips above and below the anode wire.

The CSC are placed in the first layer of the endcaps and cover the region $2 < |\eta| < 2.7$. They consist of two disks, each with eight multi-wire proportional chambers that have four CSC plates, giving four measurements for an individual track. These are faster multi-wire proportional chambers with a finer spacial resolution with respect to the MDT. There are cathodes segmented into strips with higher granularity. The CSC are arranged radially, with sections fanning out from the z-axis. One set of cathodes is arranged orthogonally to the wires for precision, and the other is parallel to the wires, providing a measurement of the transverse coordinate. This allows both coordinates to be measured from the induced charge distribution.

The gas between the strips and wires is a non-flammable mixture of Ar (30%), CO_2 (50%) and CF_4 (20%). The resolution of a chamber is 60 µm in the bending plane η and approximately 5 mm in the transverse plane.

Resistive Plate Chambers

The Resistive-Plate Chambers (RPC) are gas chambers mounted in the barrel region. In the range $|\eta| < 1.05$. They are used for muon triggering and secondary complementary coordinates, orthogonal to the precision measurements from the MDT. They are a gas based parallel electrode-plate detector formed by two parallel resistive plates held apart by insulating spacers. These are oppositely-charged anodes and cathodes. Both of these are made of a plastic laminate material with high resistivity, Bakelite. Plates used are at a spacing of 2 mm and with uniform electric field of $4.9\,\mathrm{kV\,mm^{-1}}$ across them.

The gas inside is a mixture of F_4 (94.7%), H_{10} (5%) and SF_6 (0.3%), as well as smaller percentages of C_2, H_2, and Iso-C_4. As a muon passes through the gas mixture in the electric field, a limited-ionising-avalanche multiplication takes place, centred around the primary-ionisation electron. Muons passing through the RPC ionise the gas, which sets free electrons. These are accelerated by the electric field and start ionising more of the gas atoms which leads to a chain reaction of many accelerated electrons which is called an electron avalanche.

These avalanches are formed along the ionised tracks towards the anode, which is read out by capacitive-coupling to metallic strips. The signal from ionising particles is collected by ex-

ternal aluminium strips, separated from the plates by an insulating film, with a time resolution of the order of 1 ns, a significantly shorter time scale than the separation between two bunch crossings, 25 ns. These are read out by metallic couplings on the outside of the resistive plates.

These signals are utilised as inputs for the muon triggers thanks to the rapid speed of response. They also provide a measurement of the η-ϕ coordinate of the muon, which is not measured in the MDT. A maximum of six space-points are recorded for every track.

Thin Gap Chambers

The final section of the MS are the Thin-Gap Chambers (TGC). They are multi-wire proportional chambers, similar in design to the CSC, optimised to achieve a fast signal collection. They are mounted within the endcaps to improve the muon trigger capability in the region $1.05 < |\eta| < 2.4$ and to determine the ϕ coordinate of muon trajectories in the forward direction up to $|\eta| < 2.7$.

The TGC consist of an array of anode wires that are placed between two graphite cathode layers and filled with a gas mixture. They are filled with CO textsubscript2 (55%) and n-pentane (45%) gas, with the cathode plates 2.8 mm apart. The anode wires are only 1.8 mm apart, which, along with a high electric field, leads to very good time resolution. The wire-to-cathode distance, 1.4 mm, is smaller than the wire-to-wire distance 1.8 mm. Again, a muon passing through the gas in these TGCs will ionise it and cause an electron avalanche which is collected on the nearest wire. They provide large signals and in a very narrow time window making them ideal for triggering purposes. They have a higher granularity than the RPC.

3 THE GENERATION, SIMULATION AND RECONSTRUCTION OF ATLAS DATA

If you wish to make an apple pie from scratch, you must first invent the universe.

Carl Sagan

Monte Carlo (MC) simulated processes are an integral element of particle physics analysis at ATLAS. Custom software algorithms are developed to simulate the physical processes taking place following the proton-proton collisions [107]. Simulation of both signal processes and background are critical to further understanding of physics of particle interactions and decays following proton-proton collision. Predicting the evolution of these physical processes is necessary to design meaningful analyses and validate predictions with real data taken from the detector allows us to evaluate the accuracy of our theoretical models. There are three main stages to this procedure: firstly, the generation of MC events, discussed in Section 3.1. Secondly, we simulate the interactions of particles in these events with our detector, covered in Section 3.2. This emulates the response of the real detector and the respective responses of each sub-system of the detector and is the first step to allow a direct comparison to real ATLAS data. The third stage is the reconstruction of physics objects, described in Section 3.3. Here, standardised algorithms reconstruct different types of physics objects based on signals from the detector sub-systems. The reconstruction procedure is the same for both real data and simulated events, to allow direct comparison. The specific MC samples and reconstructed objects used in the main analysis for this thesis are covered in Chapter 6.

3.1 Generation of Monte Carlo-Simulated Events

Critical to the analysis of physics within ATLAS data is the employment of MC-simulated physics processes [107]. The use of simulated collision events, generated through a multi-phased MC method allows the replication of signal processes and potential background processes. These simulations are repeatedly implemented to further anticipate predictions about event rates, kinematics, topologies and signals within the detector. The generated simulations involve the

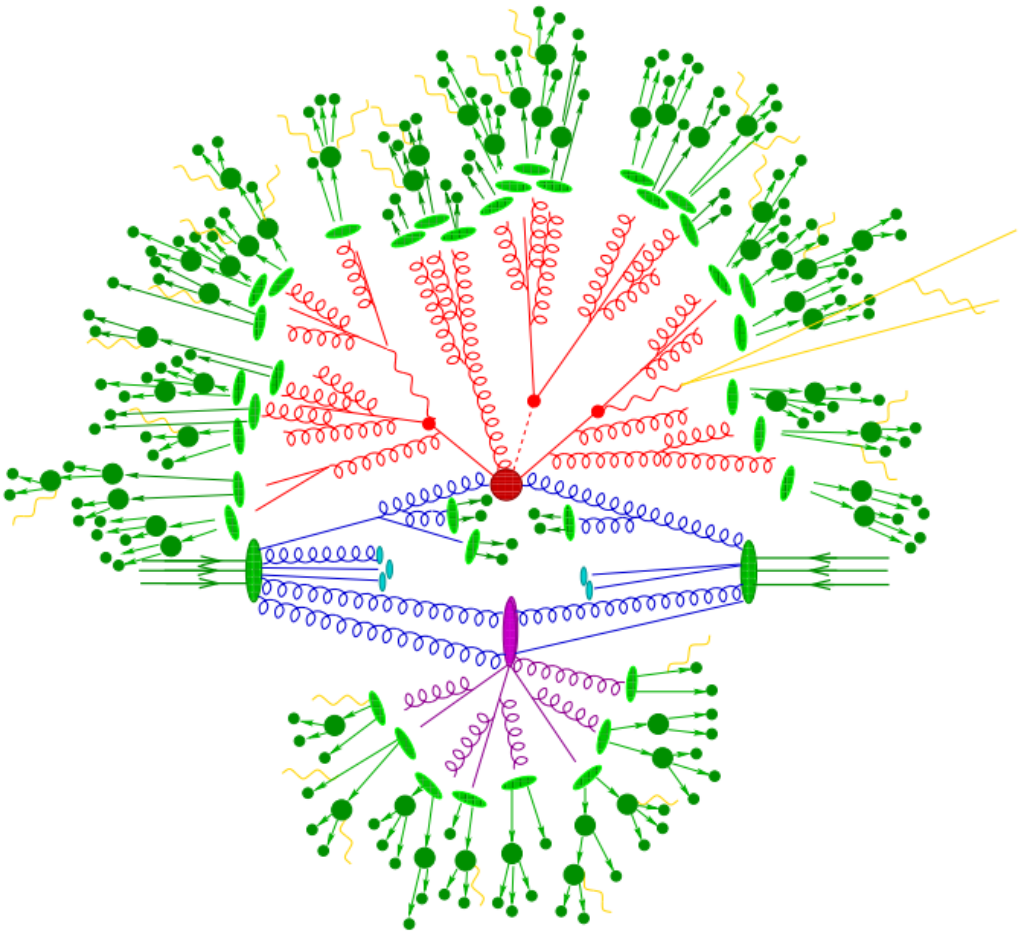


Figure 3.1: Illustration of a proton-proton collision in the LHC, as simulated by MC event generators [13]. Depicted is $t\bar{t}H \rightarrow WbWbbb \rightarrow qqb\ell\nu bbb$. The colliding protons are symbolised by the two large, green ovals. The gluons of these protons are signified by the blue, looping lines. One parton from each proton is involved in the hard-scattering event, represented by the red circle. The red, looping lines extending from the hard-scatter indicate Bremsstrahlung, as simulated by parton showers. The purple oval at the bottom and it's respective branching structure indicates a secondary scattering event, the UE. Light green ovals indicate the hadronisation of partons, and dark green ovals indicate the subsequent decays of these hadrons. Finally, the straight and undulating yellow represent leptons and soft photon radiation respectively.

use of both perturbative and non-perturbative (phenomenological) techniques to produce the events. To allow full analysis of experimental data, these generated events must then be passed through simulations of detector responses.

The execution of MC generation involves stochastic simulation of events following theoretically derived probability distributions. A target process is chosen and a selection of samples are produced through the filtering of events. Samples are divided into categories to allow the selection of relevant processes. Following this, filtering by given final states is possible; this reduces the demand on computational resources.

Descriptions of Proton-Proton collisions at the LHC depend on identification and categorisation of the Deep Inelastic Scattering (DIS) process [108]. These DIS processes are so named

due to the probing of the substructure of the each proton by the other, and the non-conservation of momentum during the collision. These processes can be hard or soft. Hard processes can be described through perturbation theory; describing soft processes, which dominate interactions at the LHC must include the use of non-perturbative QCD effects.

Most frequently, the collision includes a hard scattering process, which takes place between two constituent partons, one in each of the protons. An illustration is given in Figure 3.1. These events are defined as having either a significant transfer of momentum, large p_{T}, or a large mass scale. The target process of our analyses most often originate from these central hard scattering events and can be chosen before event generation takes place. The list of partons included within the models of our proton-proton collisions are the valance quarks within the proton, uud, the gluons that mediate strong interactions between these valence quarks, g, and all sea quarks produced through virtual $q\bar{q}$ creation originating from strong field fluctuations within the hadron.

Beyond the central hard scattering process, the collision also involves multiple softer processes, including Initial State Radiation (ISR), Final State Radiation (FSR) and the Underlying Event (UE). ISR originates from particles which are radiated by the partons pre-scatter. The ISR interacts with the hard process. FSR is emitted post-scattering. Soft scattering also occurs from additional remnants produced from the breakup of the colliding protons. All processes originating from ISR, FSR and beam remnants constitutes the UE. This approximation is limited, however, by identification and understanding of soft or collinear emissions.

Generating the events using MC simulations is broadly broken down into a number of steps involving both perturbative and phenomenological calculations. We employ a variety of specialised event generators, specialising in a range of theoretical techniques and methodologies, to suitably develop simulations of the many significant physical processes that take place within the collisions.

The first juncture to address is the modelling of the initial state of the protons within our collision. The point-like constituent partons carry a fraction of the total momentum of the proton that contains them. The momentum of our proton must be shared amongst its partons through use of mathematical processes known as Parton Distribution Function (PDF)s [82]. Following this we simulate the primary hard scattering of the partons within the proton. The simulation of the hard scattering process is divided into a Matrix Element (ME) component, involving perturbative calculations to a fixed order, and a Parton Shower (PS) component which simulates the emission of additional soft object within the event. Colour-charged particles simulated within the event are able to radiate gluons or produce $q\bar{q}$ pairs through these PS, resulting in cascades of radiation within the detector. The next stage, the simulation of Hadronisation, emulates the process by which these showering colour-charged quarks and gluons eventually reach energies low enough to form colour-neutral hadrons. Below we will review these key steps involved in producing the required MC simulated events in depth.

3.1.1 Parton Distribution Functions

The first part of the process of generating our events is through the use of Parton Distribution Functions (PDFs) [82]. During high-energy collisions, the substructure of our protons are probed. This takes place through Deep Inelastic Scattering (DIS) and involves the exchange of a boson between two partons, each from within one of the two protons in our collision, transferring a four-momentum of modulus q. We cannot know the flavour and momentum of partons within our colliding protons, as this calculation is non-perturbative. This is because low energy, soft QCD effects dominate the parton interactions. PDFs are therefore used to represent the probability of specific parton flavours found within the proton and the probabilistic momentum distribution these constituent partons, i, possess. PDFs can be seen as a description of the change of the structure functions of the proton's constituents as a function of the running strong coupling, α_S. They are used throughout the stages of the MC simulation; the ME, PS and UE.

PDFs rely on the momentum transfer between interacting partons, Q^2 through exchange of bosons with momentum $-q^2$ and the fraction of the total proton's longitudinal momentum, x_i. This is known as the "Bjorken-x" [109]. The structure of the proton is represented as $f_i\left(x_i, Q^2\right)$. Factorisation theorem declares that the differential cross-sections, $d\sigma$, of a given hard process can be given as [110]

$$d\sigma = \sum_{I,j} \int_0^1 dx_i \int_0^1 dx_j \; d\sigma_{i,j}^{part} f_i\left(x_i, Q^2\right) f_j\left(x_j, Q^2\right) \tag{3.1}$$

Here $d\sigma_{i,j}^{part}$ represents the differential cross-section of the hard process between the pair of partons. PDFs are universally applicable and do not depend on the hard process being considered, as they are functions of the momentum fraction.

The source of these PDF shapes are acquired from fitting measurements to experimental measurement data from both fixed-target and collider experiments, including the Tevatron and HERA. This allows the calculation of the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi equations (DGLAP) evolution equations, which are used to compute the dependence of the PDFs as a function of Q^2 for a particular parton [111, 112, 113]. This means that the measured PDF can be transferred to experiments within much higher regions of Q^2, such as the LHC.

The accuracy of the PDF is variable, it can sometimes extend up to Next-to-Next-to-Leading Order (NNLO), but it more often set at Leading Order (LO) or Next-to-Leading Order (NLO). A decision on the flavour scheme used within the calculation also must be made, most often between the four-flavour (4F) and five-flavour (5F) schemes. The 4F scheme implements massive b-quarks within the calculations, such that the constituents of the proton are restricted to the four lightest quarks and the gluon. b-quarks are, however, included within the final state. In the 5F scheme b-quarks are factored in as massless partons, and are therefore also included in the initial state.

3.1.2 Matrix Element Calculations

The second part of the event generation involves the computation of the hard scatter within the event, describing the interaction of partons within the proton, and the outgoing particles they produce. This is done through use of the Matrix Element (ME) calculation. Matrix Element (ME) calculations involve the perturbative quantum field theory calculations of the Feynman diagrams of processes of interest within the event, and the computation of the partonic cross-section $\hat{\sigma}_n$.

Hard processes involve a momentum transfer at large scales, $Q^2 > \mathcal{O}(1\,\text{GeV})$. At these energies the running coupling of QCD, α_S, drops down to ≈ 0.1, where perturbation theory can be used. These processes can therefore be constructed from first principles using the ME of the interaction, simulated through the use of QFT techniques. PDFs simulate the incoming partons using the ME in an expansion of α_S. This leads to the calculation of a probabilistic distribution of the partons outgoing from the interaction. All hard processes are considered in the simulation, including hard emissions of energetic quarks and gluons, either through a quark radiating a gluon ($q \to qg$), or when a gluon either decays into a quark-antiquark pair ($g \to q\bar{q}$) or gluon pair ($g \to gg$). There are dedicated event generators to supply higher order calculations.

Associated scattering matrices can be calculated to differing orders. Many of these processes are now determined to NLO precision in perturbation theory expansion of α_S, and can be normalised to NNLO. Multi-purpose event generators will provide a comprehensive list of LO matrix elements. MADGRAPH [114] and POWHEG [115] generators are commonly used to compute these Matrix Element (ME) calculations to NLO precision.

The use of ME calculations in this process cancels out Ultraviolet (UV) divergences in higher-order computations; however, if larger multiplicities of final state objects are included, these cancellations are nullified: contributions from soft collinear emissions bring new divergences. These contributions are removed through an introduced cut-off, such that extraneous soft objects are excluded. These are consequently added back during the Parton Shower (PS) phase of the simulation. Any overlap with the PS is removed through a matching procedure, determining the separation in phase spaces covered by the ME and PS.

3.1.3 Parton Showers

The third phase in event simulation is the Parton Shower (PS) phenomenological modelling. The PS builds on the ME describing the evolution of partons preceding and following the hard scatter. Due to the lower energies involved, the emission of the soft objects cannot be modelled with ME calculations. Calculations are non-perturbative, therefore differing methods must be used. Consequently, higher-order calculations can be taken into account at this stage, unlike during the ME calculations. These higher-order contributions are additional, soft emissions from QED and QCD processes. Determining these higher-order calculations, however, requires approximations in which the dominant contributions within each order are the only ones included.

Phenomenological models use algorithms based on step-wise Markov chains [116]. These algorithms give probabilistic decisions on soft-processes, including gluon emission by a quark ($q \to gq$), or pair production from either a quark-antiquark pair ($g \to q\bar{q}$) or a gluon pair ($g \to gg$). At each step of the algorithm a decision is made as to whether these processes will occur. As the parton energies decrease they undergo Hadronisation.

During intermediate levels of momentum transfer (Q^2) there is potential for crossover between hard emission to be treated peturbatively, and softer processes to be treated as part of the PS, leaving the potential that double-counting might take place. Use of the Catani-Krauss-Kuhn-Webber (CKKW) [117] and Michelangelo L. Mangano (MLM) [118] algorithms are utilised to identify the domain of these processes and assign them to be treated either through ME Calculations or within the PS in a way that avoids significant discontinuities in observable spectra.

The most commonly used PS models are developed by the PYTHA [119], SHERPA [120] and HERWIG [121] collaborations.

3.1.4 Hadronisation

Following the PS, the simulated partons reach a value of momentum transfer (Q^2) approaching the order of $\Lambda_{QCD} \approx 200\,\text{MeV}$. At these energy scales the colour confinement property of QCD becomes relevant to calculations, as all evolved soft partons (with exception of the top quark) begin to Hadronise. The Hadronisation of these partons into colour-neutral hadrons is regulated by non-perturbative QCD, meaning specialised phenomenological modelling techniques are once again required. Unstable hadrons created within this process will further decay into stable final-state objects. This is also simulated during this phase. This requires the use of non-trivial parameterisation decisions to simulate the decay chains of these unstable hadrons.

The two models most commonly used for modelling these processes are the cluster model, by HERWIG [122], and the Lund string model, by PYTHA [123]. Both of these methods contain a high number of parameters, they must therefore make use of empirical methods, fine-tuned to experimental data [124, 125]. The final state of the Hadronisation phase then consists of objects that are stable on the timescale of colliders, and therefore can be passed onwards to a simulation of the detector.

3.1.5 Underlying Event and pile-up

The final phase of the simulated event is the modelling of the Underlying Event (UE) [126]. Alongside the hard scattering processes, involving only two quarks or gluons, interactions from extra partons can produce extra soft interactions at lower scales of energy. LHC processes can be regarded as Multiple Parton Interactions (MPI), generating multiple instances of underlying hadronic activity. These are referred to as the UE.

Pile-up from non-primary interactions must also be considered. There are two categories of pile-up to be considered: in-time and out-of-time pile-up. For in-time pile-up the source of

interactions are collisions between proton pairs of the same bunch crossing as the hard scatter. Out-of-time pile-up occurs due to collisions in a different bunch crossing. This pile-up is modelled and overlaid on the MC simulation using PYTHA.

The low energy scales of these interactions means, once again, that specialised phenomenological models are used. These are reliant on parameters tuned to data. The resultant Hadronisation of these interactions must also be taken into account, and are simulated using related models.

3.2 Detector Simulation

Following the event simulation, we are left with a series of truth level events: a list of four-vectors representing particles defined as "stable"- with lifetimes of approximately 3×10^{-11} s- and holding the kinematic information of the event. These truth level events depict the physics objects prior to their interaction with the detector, meaning that for meaningful comparison with data from the experiment this simulated data must be passed through another simulation, this one replicating the effects of the detector and the subsequent signals from the detector sub-systems [127].

To produce the detector simulation we use GEANT4 [128], a software toolkit integrated within ATLAS offline software. The result of this is an output in an identical format to experimental data passing through the ATLAS TDAQ. This means that both our MC simulation and real ATLAS data can be processed in a consistent way through the same reconstruction software. There is also use of faster simulations, such as AltFastII (AF2) [129] , using a parameterised simulation of detector response.

3.2.1 Propagation

The first stage of simulating the effects of the detector involves recreating the propagation of MC simulated physics objects. This simulates the interactions between these particles and the numerous sub-systems integrated within the detector. The interactions must be equivalent to the conditions experienced by real particles undergoing detection, including undergoing intermediate particle decays and scattering. This stage also accordingly simulates the response of the detector, including ionisation and radiation. "Hit" files are produced, containing simulated track positions and energy deposits in the sub-detectors, and are passed onto the following stage: "Digitisation".

3.2.2 Digitisation

The second stage of our detector simulation, "Digitisation", transforms our "Hits" through a reproduction of the electrical response of our detector subsystems. Therefore, the tracks and energy deposits in our simulation are translated into digitised values for output-style electronic

signals: currents, voltages and associated times. During this process a pre-generated model of background noise found within the detector is inserted. Incorporating the background at this stage allows us to save CPU processing time by preventing repeated simulation of information independent from the event. The noise consists of a number of distinct items: Extraneous soft collisions, such as pile-up interactions; cavern background, including low energy neutron physics; beam halo; and cosmic rays.

This MC simulation data is then output into a format known as a Raw Data Object (RDO), the same output format as given by the ATLAS TDAQ system. Having our MC simulation output contain equivalent information to our real experimental data allows processing by the same trigger and object reconstruction software, such that the simulations and real data can be explicitly compared.

Additionally to the RDOs we produce Simulated Data Object (SDO) files. These represent a "truth" level record of the event, with information on the true identity of particles at each of the vertices and all corresponding tracking and decay information. These truth level objects can (and will) later be matched to reconstructed objects, allowing comparison and evaluation of algorithms used in reconstruction processes by identification of potential misreconstructions.

3.2.3 Fast Simulation

The complete procedure of running the detector simulation involves an incredibly detailed account of each process taking place within every subsystem and is labelled as the "Full Simulation" (fullsim for short). Simulating a single event using fullsim typically takes several minutes. When requirements dictate the need for millions of events for a given sample this timeframe can be significant and a quicker option may be preferable. Thankfully, there is a stream-lined option available for detector simulation: the "Fast Simulation" (fastsim).

Employing fastsim can reduce the simulation time by a full order of magnitude – particularly useful for large dataset requirements – and can reduce the strain on computing resources. ATLAS uses the software AltFastII (AF2) [129] to generate these fastsims. AF2 utilises the full GEANT4 process software [128] for the ID and MS simulations, and uses FastCaloSim for the calorimeters.

The increase in speed of AF2 comes from parameterising the simulation of our physics particles energy responses and distribution within the ATLAS detector, simplifying the PS process which would normally take up >90 % of the processing time, now reduced to a few seconds per event. Nevertheless, this fastsim, however useful in reducing processing times, must be used appropriately: analyses requiring in-depth calorimeter information, such as JSS, necessitates that we use the full PS to avoid significantly less accurate models. This is the case in this analysis. JSS will be discussed in Section4.5.

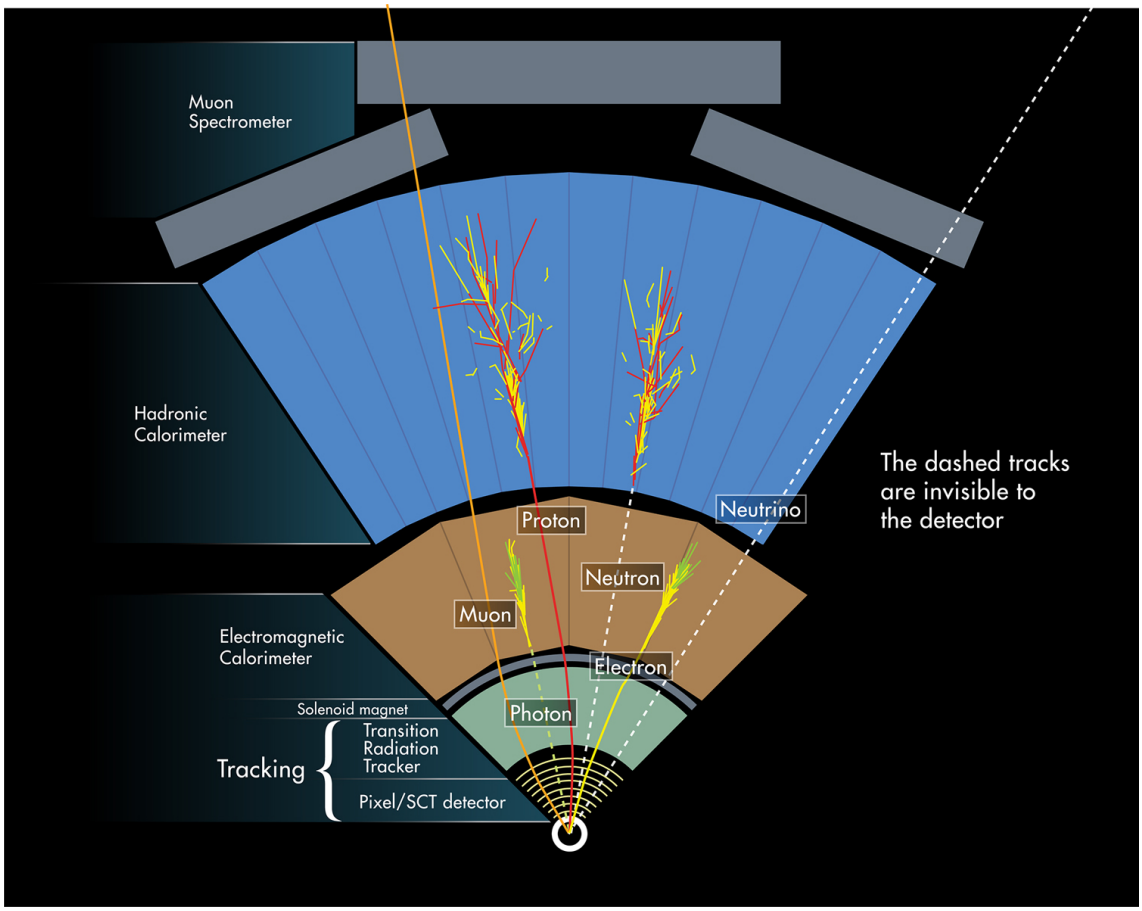


Figure 3.2: Diagram representing the detectable signatures of different physics objects within the ATLAS subdetectors [14]

3.3 Object Reconstruction

Following, the collection of real data from the ATLAS detector, we are left with data that is ready to be processed into physics objects for identification and analysis. The data from our RDOs, containing digitised detector signals and responses, can now be converted back into tracks and calorimeter information, which in turn are reconstructed into physics objects through the use of reconstruction algorithms. Examples of physics objects are shown in Figure 3.2.

In the case of real detector data, preliminary selections on events are executed by the ATLAS TDAQ system: relevant signals of the desired physics objects are pinpointed by software algorithms running in the HLT (described in detail in Section 5.1.2. These algorithms are required to run quickly; however, once these events are actually recorded we can use more rigorous reconstruction algorithms to analyse the data.

Object reconstruction at ATLAS is executed through the Athena software framework [130] to combine information from the various subdetectors to a specification depending on the objects used and the analysis objectives. This data is then output in the format of an Analysis Objects Data (AOD) for physics analysis. In the first instance, physics objects are reconstructed with a broad set of requirements. Analysis-dependent specifications allow, and dictate, tighter

object definitions to increase object selection purity at the detriment to broader efficiency.

A range of reconstructable objects exists: Tracks and Vertices, Electrons and Photons, Muons, taus, $E_{\mathrm{T}}^{\mathrm{miss}}$ and Hadronic Jets. The final major physics objects to be reconstructed within our simulations, Hadronic Jets, will be discussed in full detail in Chapter 4.

3.3.1 Tracks and Vertices

The first set of object reconstructions we shall look at is that of the tracks and vertices, both based on data coming from the ID. Tracks are the reconstruction of pathways taken by charged-particles through the ID [131, 132]. Vertices are defined as the positions at which single particles decay, or pairs of particles interact [133]. Tracks and Vertices are first reconstructed themselves, before being matched to separate physics objects for further uses. Tracks are vital for charged-particle reconstruction, and vertices are required both for flavour tagging in jets and for quantifying pile-up. A "Primary Vertex" is defined as the vertex at which the hardest event scattering is taking place.

Tracks

As charged particles make their way through the ID and pass the different layers within (the pixel layer, SCT and TRT) they generate hits. These hits are combined through a number of algorithms and reconstructed into tracks [131, 132, 133]. There are three main algorithms which are used for track reconstruction. The primary track-reconstruction algorithm uses what is known as an "inside-out" approach: starting from the innermost layer, requiring at least 3 grouped, consistent hits in the inner pixel and SCT layers, with each of the hits required to pass a minimum p_{T} requirement of >1 GeV. Subsequently the track reconstruction moves to extrapolate the tracks to include compatible hits from the TRT through use of a combinatorial Kalman filter [134]. A secondary "outside-in" algorithm, aiming to target later decays of neutral particles within the ID, is also used. Known as back-tracking, it is seeded within the outermost layer, the TRT, and works its way inwards towards the pixel layer, considering hits not utilised in the "inside-out" reconstruction. Thirdly, a final track reconstruction algorithm using only hits from the TRT is used. Following the initial steps of reconstruction, selections can be applied to reach the appropriate track quality for analyses.

Charged particles within the ID undergo acceleration from the magnetic field, and as a result the pathways they leave are helicoidal, with curvature inversely proportional to the p_{T} of the particle. The tracks can be parameterised through use of five track perigee parameters, all given relative to the origin of the ATLAS co-ordinate system and measured at the point closest to the z-axis. Firstly, the combination of two space co-ordinates: pseudorapidity, η and the azimuthal angle, ϕ. The next are impact parameters: the transverse impact parameter, d_0, the distance of closest approach between the track and origin, measured in the transverse plane; and the longitudinal, z_0, the distance measured in the longitudinal direction. Finally, the p_{T}:

which is proportional to the magnitude of the magnetic field, B, and the bending radius of the trajectory, R (see Equation 2.7.

Vertices

Vertex reconstruction at ATLAS is required to identify physics objects produced in the correct hard scattering event from those produced in other pile-up collisions [133]. Proton-proton collisions are distributed along the z-axis of our beamline, in coordination with the finite length of the proton bunches. To correctly ascertain the collision of origin of each vertex, we ensure the tracks that constitute it are traced back to the beamline.

For a track to be defined as coming from a Primary Vertex (PV) it must lie within the beam-spot area, the area surrounding the collision point where the two proton beams overlap, and have at least two selection-passing tracks associated to it. A higher minimum number of tracks can be determined to decrease the likelihood of events additional to our event of focus degrading the vertex reconstruction.

Reconstruction of vertices works on the principle of grouping reconstructed tracks back to a common interaction point using an adaptive vertex-fitting algorithm [135, 136]. Each track used in the reconstruction of vertices must pass a minimum p_{T}>400 MeV and have a suitably central pseudorapidity, η<2.5.

Vertex reconstruction is an iterative process. Reconstruction ends when no additional vertices can be constructed or if all possible tracks are associated to vertices. To begin our vertex construction we first identify a seed, signified by a global maximum of reconstructed tracks within the z-direction. The centre of the vertex is decided by fitting tracks using the least squares fitting method. Outlying tracks, far from the vertex centre, are more likely to be noise tracks, not originating from the vertex we seek to reconstruct, and are therefore assigned lower weights in the fitting process.

The true position of the vertex centre, and the identity of any real outlier tracks, are unknown to us. Each iteration of track-to-vertex matching results in the weights of outlier tracks being diminished. Each iteration of the reconstruction increases the accuracy of vertex construction and identification of outlying tracks. Once the vertex centre ceases to change between iterations, the algorithm appoints the vertex complete, and moves to find a new seed, using those tracks not associated with the current vertex. This continues until all tracks are clustered into vertices, or no further vertices can be reconstructed.

Complications through pile-up interactions can lead to multiple vertices being successfully identified. The assignment of a true PV, representing the main hard scatter of our event, is given to the vertex with the highest sum of squared associated track momenta ($\sum p_T^2$). All other PVs found are, therefore, assigned as pile-up vertices. The number of these pile-up PVs within the event, Number of Primary Vertices (NPV), is related to the average number of interactions per bunch crossing, $\langle\mu\rangle$, used as a direct measurement of the pile-up for calibrations.

We also define a Secondary Vertex (SV) as those which lie outside of the beamspot region.

These SVs originate from tracks displaced from the PV, indicating the particles with lifetimes sufficiently long to traverse a measurable distance before decaying (typically mms). This feature can aid in the flavour-tagging of jets, by identifying the presence of these particles.

3.3.2 Electrons and Photons

Within ATLAS both electrons and photons are reconstructed using a combination of information from the ECAL and ID [137]. The experimental signature characteristic of photons and electrons within the ATLAS detector is an electromagnetic shower taking place within the ECAL. Electrons also have an associated track within the ID. For electrons and photons, these are reconstructed within the central region corresponding to the acceptance of the ID and the highly segmented section of the ECAL, $|\eta| < 2.47$.

The first step to reconstructing electrons and photons involves the clustering of calorimeter energy deposits from seeds, which begins with identifying seed-clusters. These seeds are energy deposits within the ECAL that have p_{T} >2.5 GeV, when the energy of all layers of the calorimeter are added together in a "tower" structure. Any duplicates are removed following energy comparisons with nearby seeds. A standard threshold is chosen to optimise the efficiency of the reconstruction, which helps to minimise any possible contribution from noise originating from electronic or pile-up sources.

A method commonly used in the reconstruction of these physics objects is the clustering by sliding-window algorithm [138, 139]. This process searches for clusters by defining a fixed longitudinal tower within the ECAL as a window, scanning over blocks of three-by-five in units of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, which corresponds to the granularity within the middle sampling layer of the ECAL. Towers with the most energy deposited are identified, and matched to tracks within the ID.

Track matching starts with searches for those with p_{T} >500 MeV. These tracks are extrapolated from their last measured point inside the ID up to the middle sampling layer of the ECAL. The position within $\eta - \phi$ space extracted from this extrapolation is them compared with our seed cluster, and the two are deemed geometrically matched if $|\Delta\eta| < 0.05$ and $|\Delta\phi| < 0.1$. If more than one track meets these criteria, however, the track with the smallest geometric distance, R, is chosen.

If there are multiple possible tracks within the ID then we must identify the primary track in order to determine the kinematics and charge of the electron. Any tracks that have hits in the pixel or SCT layers are prioritised, and the one closest to the centre of the cluster is chosen. We then define the object based on the number and location of the tracks found: the object is labelled as an electron candidate when at least one track is found, or as a photon if there are no tracks, or tracks are only found within the TRT.

It is also possible to reconstruct converted photons that produce an electron-positron pair through interaction with materials within the detector. This process is characterised by the presence of two oppositely-charged tracks that are matched to the seed cluster. If the opening

angle of the electron-positron pair is large, however, a single track with no hits in the innermost silicon layers may be found [140].

Upon a successful matching between a seed cluster and a track, the cluster window is enlarged to three-by-seven units within the barrel region ($|\eta| < 1.475$) and five-by-five units in the end-caps ($1.375 < |\eta| < 3.2$), centred on the original seed cluster centre. These new, larger cluster windows are used to determine the energy of the electron candidate. Four separate contributions are collated to determine the total energy of the cluster. The first is the estimated energy deposit in material in front of the ECAL shower sub-detector, this is used to measure any energy lost by the particle before it reaches the ECAL. The second is the measured energy deposit within the found cluster itself. The third contribution comes from energy deposits within the ECAL that lie outside of the cluster itself: the lateral leakage. The final contribution comes from estimated energy deposited beyond the ECAL, and is called longitudinal leakage. For the final energy of the electrons we also combine the cluster energy with any track momenta.

Following reconstruction of electrons and photons, identification algorithms are employed to exclude any misidentifications. "Fake" electrons signals can come from other objects, including muons, photons, jets, pion decays and other hadrons. To identify fakes a likelihood discriminant is utilised, gathering information from ID tracking, the calorimeters, combined track-cluster variables, such as shower shape, and radiation deposited within the TRT.

The reliability of real electron reconstruction can be improved by including additional cut-based conditions to provide a better level of separation between real and fake electrons. There are three different WPs defined for electron identification: loose, medium and tight [141]. These WPs possess increasingly improved rejection of background (therefore an increased sample purity) through a process of tightening criteria on variables, but has the downside of lowering the levels of identification efficiency. The signal efficiency in electron identification is given by the ratio of electrons passing identification criteria to the total number of electron candidates. For photons there exists only two WPs: loose and tight. Both the shower shape and the amount of hadronic activity taking place behind the EM cluster lie at the centre of photon identification. EM showers caused by photons and those caused instead by neutral mesons can be distinguished from one-another through the investigation of energy deposits within the first two layers of the ECAL [140].

3.3.3 Muon Reconstruction and Identification

Reconstruction of muons within the ATLAS detector involves the use independently identified tracks from both the ID and MS [142, 143]. These muon-track candidates are constructed through seed segments from within the middle layers of the MS, where more trigger hits are available. These seed segments are combined through a matching process with hits located in segments from both the outer and inner layers of the MS. For a segment to be considered for this reconstruction all tubes that it covers must contain a "hit". We require at least two of these segments to construct a track.

There are four different types of muon reconstruction algorithm, all within different requirements [144]. In the first of these, a "combined" process, the track reconstruction is performed independently in both the ID and MS before a combined track is prepared. This combined track is fashioned with a global fit, a process that requires considerations of energy lost through the calorimeter. The process most commonly used to produce this is an "outside-in" pattern recognition algorithm. Firstly the reconstruction takes place within the MS and is then extrapolated inwards to match tracks of comparable p_{T} that are located within the ID. This provides a good momentum resolution for all ranges of muon p_{T}.

The second muon reconstruction tactic is the "Segment Tagged" process. In this procedure muons are reconstructed via the matching of a singular MS segment track, located within the inner layer, to a track from within the ID. This technique is most applicable to muons with low p_{T}, or those that pass through MS regions with low acceptance rates.

The third method is the "Calorimeter-Tagged" approach. This relates to muon candidates with ID tracks that can be matched to energy deposits within the calorimeter that signal the presence of a Minimum Ionising Particle (MIP). This MIP subsequently leaves no tracks within the MS. This process is used to recover acceptance for muons within the central region of the detector, $|\eta| < 0.1$, that may have only been partly recorded due to a crack within the barrel of the MS.

The fourth and final algorithm used to reconstruct muons is the "Extrapolated" method. Here a track from within the MS will signify a muon with a trajectory that lies inside the acceptance range of the MS, but outside of the ID acceptance, $2.5 < |\eta| < 2.7$. The trajectory of the MS track is extrapolated backwards and is required to have a loose compatibility with the interaction point on the beam-line at the centre of the detector.

Muon candidates within ATLAS can be non-prompt, meaning that they originate from decays of charged pions and kaons, instead of from the primary vertex of the event. Further identification restrictions are applied in order to separate these from prompt muons. One of these requirements is the track quality standard within both the ID and MS sub-detectors. Another is criteria placed on the compatibility of charge and momentum between the ID and MS. A normalised χ^2 distribution of the combined MS-ID track fit is also used.

As with electron reconstruction, three WPs are used in muon identification: loose, medium and tight.

3.3.4 Tau Reconstruction and Identification

Tau identification at ATLAS is a more complicated process [145, 146]. This is due to the heavy mass of the tau leading to a short lifetime, meaning they decay within the beam-pipe region before reaching our detectors. 32.5 % of tau decays are leptonic, and are indistinguishable from light leptons originating from our event. The remaining 64.7 % of tau decays are hadronic, and therefore the identification of these taus rely on reconstruction of their decay products, hadronic jets. See 4.

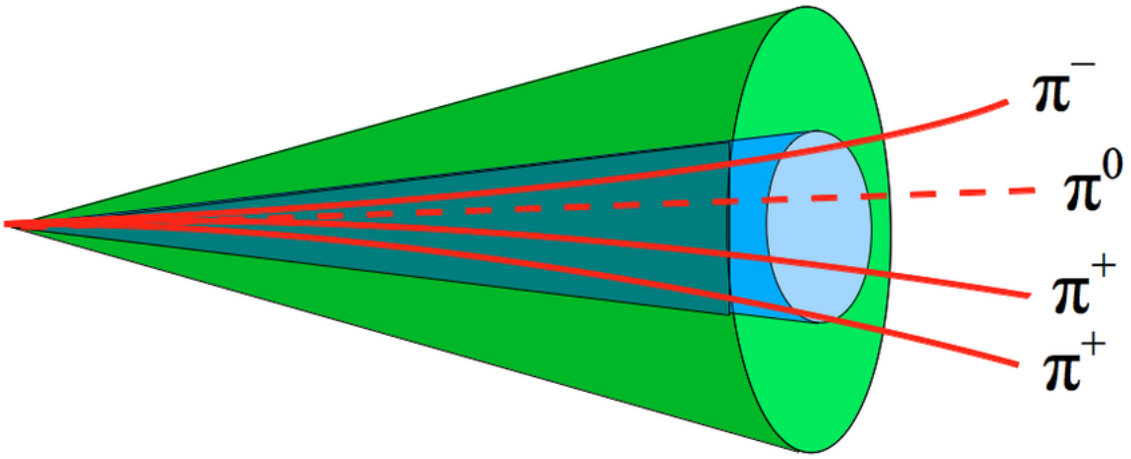


Figure 3.3: Illustration of a 3-pronged hadronic decay of a τ -lepton [15]. The three charged hadrons produced in the decay leave charged tracks within the ATLAS ID that can be used to classify the hadronic decay.

More than 90 % of hadronic tau decays that take place, do so though just five dominant decays. These five decays modes leave either one or three charged hadrons, up to two neutral pions (π^0) and a tau neutrino. An example of a hadronic tau decay with three charged hadrons is shown in Figure 3.3. To identify hadronically decaying taus, a combination of shower shape variables and tracking information is used in a multivariate algorithm using a Boosted Decisions Tree (BDT). These taus are required to have min $p_T > 15\,\text{GeV}$, be within $|\eta| < 2.5$ and have either one or three associated tracks. Hadronic tau reconstruction has an efficiency of 55 % for one charged decay products, or 40 % for three, with a rate of fake quark- and gluon-initiated jets misidentification of below 1 % [146].

3.3.5 Missing Transverse Energy

Missing transverse energy ($E_{\text{T}}^{\text{miss}}$), can be defined as the imbalance of visible transverse momentum within the ATLAS detector in a given event. The hermetic coverage of the ATLAS detector allows us to infer from $E_{\text{T}}^{\text{miss}}$ the presence of undetected particles. These could be particles that are known to be invisible to ATLAS, such as neutrinos, or could include potential Dark Matter (DM) candidates [147] or evidense of physics beyond the SM, such as Supersymmetry (SUSY) [148]. When these particles are produced in our events, there will be a tell-tale, significant momentum imbalance within the plane transverse to the direction of the collision [149, 150].

As we know that momentum in directions transverse to the z-direction of the collision must be conserved, this missing transverse momentum ($\mathbf{p}_{\text{T}}^{\text{miss}}$) can be calculated as the negative vector-sum of the four-momenta of all visible detected physics objects in this event:

$$\mathbf{p}_{\text{T}}^{\text{miss}} = -\sum_i \mathbf{p}_{\text{T}}^{\text{i,obj}} \tag{3.2}$$

Where $\mathbf{p}_{\text{T}}^{\text{i,obj}}$ represents the transverse momentum of the i -th visible detected physics object in the given event. The $E_{\text{T}}^{\text{miss}}$ is equal to the magnitude of $\mathbf{p}_{\text{T}}^{\text{miss}}$.

Calculating the full E^{miss}, as opposed to only that in the transverse plane, would be impossible as the longitudinal energy of the initial partons in our collision are unknown and vary with each event. We can, however, make the assumption that these partons have a negligible momentum in the transverse direction and can expect that the net transverse momentum of all detected objects is balanced, unless particles are missed by our detector.

JETS AND JET RECONSTRUCTION

4

Nothing exists except atoms and empty space; everything else is opinion

Democritus

The final major objects requiring reconstruction within ATLAS are Jets. A full understanding of the physics in ATLAS necessitates the study colour-charged particles created within the collisions at the LHC. However, Colour Confinement complicates this, due to the hadronisation of these particles. The hadronisation of quarks and gluons leads to a collimated stream of showering, energetic hadrons [151]. The pattern of energy deposits they leave within the detector, and in particular the calorimeter, are reconstructed into four-momentum vectors called jets. Understanding the topology of these signals are vital to reconstructing the physics of these jets. To replicate the physics from the original event we must endeavour to collate these energy deposits, from the calorimeter or otherwise, into sets originating from common origins. For these processes to be successful required agreement on certain sets of definitions must be used within Jet Reconstruction, whether it be the inputs used – definitions of calorimeter signals and calibrations, reconstructed tracks, generated particles and partons within MC samples – or the reconstruction algorithms and treatment processes applied upon them. These inputs are discussed in Section 4.1 and the clustering methods used to reconstruct our jets are covered in Section 4.2.

Jet construction methods must aim to build jets in an increasingly complete, reliable and accurate way. Jet reconstruction must be consistent and methodical, so as to meet theoretical and experimental requirements, therefore inputs, both from our detector and from out MC simulations, must be stringently defined. Following reconstruction, we can use these Jets as proxies for the partons they originate from. This means strict requirements on reconstructed jets must be placed, so that the most precise and accurate information can from gained them. Treatment processes are also applied, including Jet Cleaning and standardised Jet Calibrations, discussed in Section 4.3.

Through the process of reconstructing Jets at ATLAS, the aim is to re-establish the correl-

ations between the sources of these jets, allowing discovery of the physical processes taking place within the event. It may also allow the uncovering of the original particles coming from the hard scattering. Flavour Tagging involves classifying characteristics of jets originating from defined origins. The current state of this is explained in Section 4.4. There are limitations to these methods; however the use of JSS techniques provides further opportunities for jet tagging. See Section 4.5.

Some energy within jets does not originate from partons coming from the hard scatter, and instead come from external sources, such as the UE or pile-up. Techniques that examine JSS in order to algorithmically remove these extraneous contributions, known as Jet Grooming, are covered examined in Section 4.6. Shortcoming in current Jet Grooming techniques require development and testing of novel methods: soft drop grooming (soft drop grooming) is a Jet Grooming algorithm that is able to remove contributions from Non-Global Logarithms (NGL) in an Infrared Safe (IRS) manner. This is reviewed in Section 4.7.

4.1 Jet Inputs

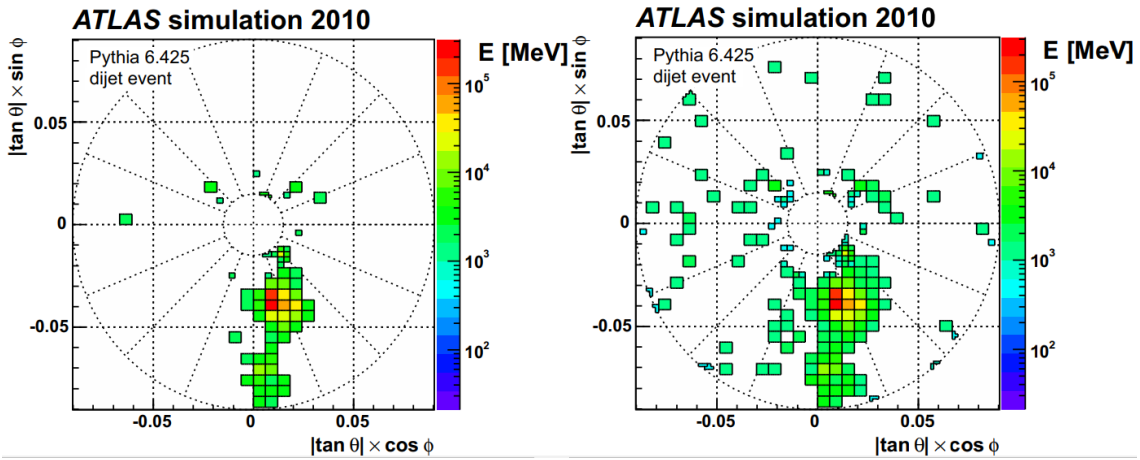
Firstly, we will address the inputs required to build our jets. From MC simulated events we have access to "Truth Particles": stable, final-state particles allowing reconstruction of "Truth Jets" that represent the physics of the event before interaction with our detector takes place. This is not possible for real data, so we therefore form our reconstructed jets using data both from the energy deposited within the calorimeter and from the tracks from the ID.

4.1.1 Topoclusters

To start the Jet Reconstruction process, we must gather the components of our jets: *topoclusters* [16]. Jet clustering methods rely on topological clusters of cells within our Hadronic Calorimeter (HCAL). We can search for these energy deposits within these calorimeter cells to start our process: however, we must bear some things in mind. Firstly, that the calorimeter cells within our detector are exposed to a base level of "noise": information coming from sources beside the hadronic jet we are attempting to analyse, or even external to the physics of our hard scattering. This can come from detector, electronics and from the UE/pile-up. We can estimate the average noise for our events through a combination of electronic noise and pile-up:

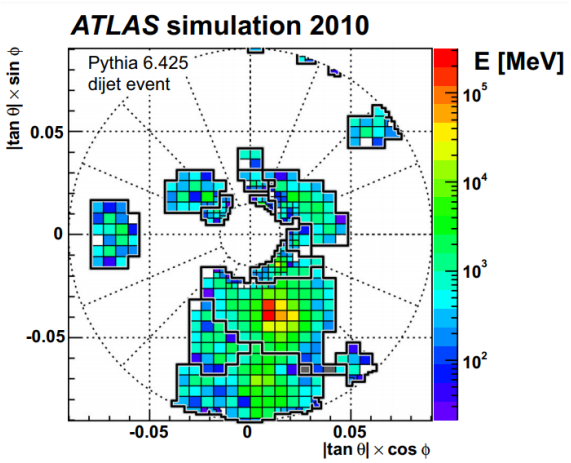
$$\langle \sigma_{\text{noise}} \rangle = \sqrt{\sigma_{\text{noise}}^{\text{pile-up}^2} + \sigma_{\text{noise}}^{\text{electronic}^2}} \tag{4.1}$$

To allow efficient data reconstruction we must adjust this average for corrections. The changing nature of pile-up, linked to the differences of instantaneous luminosity throughout any given run of the LHC, means that this average cannot accurately represent the pile-up at one any point in time: for much of the spectrum of pile-up the average is not optimal. This equation assumes a predicted level of pile-up, based on the planned luminosity and that this noise will follow a Gaussian distribution.



(a) Seed cells

(b) Seed cells with growth cells



(c) Final topoclusters

Figure 4.1: Stages of topocluster formation in the ATLAS calorimemter for a MC simulated event. Shown in (a) are cells used to seed the topoclusters. (b) shows the inclusion of topoclusters growth cells. Finally, (c) shows the completed topoclusters with boundary cells. [16]

To overcome the issues caused by this noise within calorimeter cells we can set an energy threshold. We set our threshold energy as four times the average expected noise within a given cell:

$$\left|E_{cell}^{seed}\right| > 4 \times \sigma_{noise} \tag{4.2}$$

Following this, we can search for cells with energy exceeding this threshold to seed our jet reconstruction, forming proto-topoclusters, and build the remainder of our topoclusters from neighbouring cells meeting lower energy thresholds. This stage is shown in Figure 4.1(a). We next inspect the energy within neighbouring calorimeter cells, directly adjacent to our seed either within the same layer, or into the next layer. We search for neighbouring cells with an energy surpassing the threshold of twice the expected noise:

$$\left|E_{cell}^{growth}\right| > 2 \times \sigma_{noise} \tag{4.3}$$

These so-called growth cells are then added to our proto-cluster. This is shown in Figure 4.1(b). If two neighbouring seed cells are found they are merged into a single proto-cluster. If more

than one seed cell neighbours the same growth cell, they are all merged into a single seed. Following this, we look once more for a set of cells, this time directly adjacent to our growth cells and with no energy threshold. These are boundary cells:

$$\left|E_{cell}^{boundary}\right| > 0 \tag{4.4}$$

The addition of the boundary cells completes the process of building our topoclusters. This stage is illustrated in Figure 4.1(c). All topoclusters have mass set as zero, meaning the energy of each topoclusters is equal to the momentum. Each of these topoclusters should ideally be a proxy for a single particle from within our event, however merging of topoclusters can be caused by multiple particles in close proximity. To overcome this we can additionally use tracking information from the ID to identify separate particles.

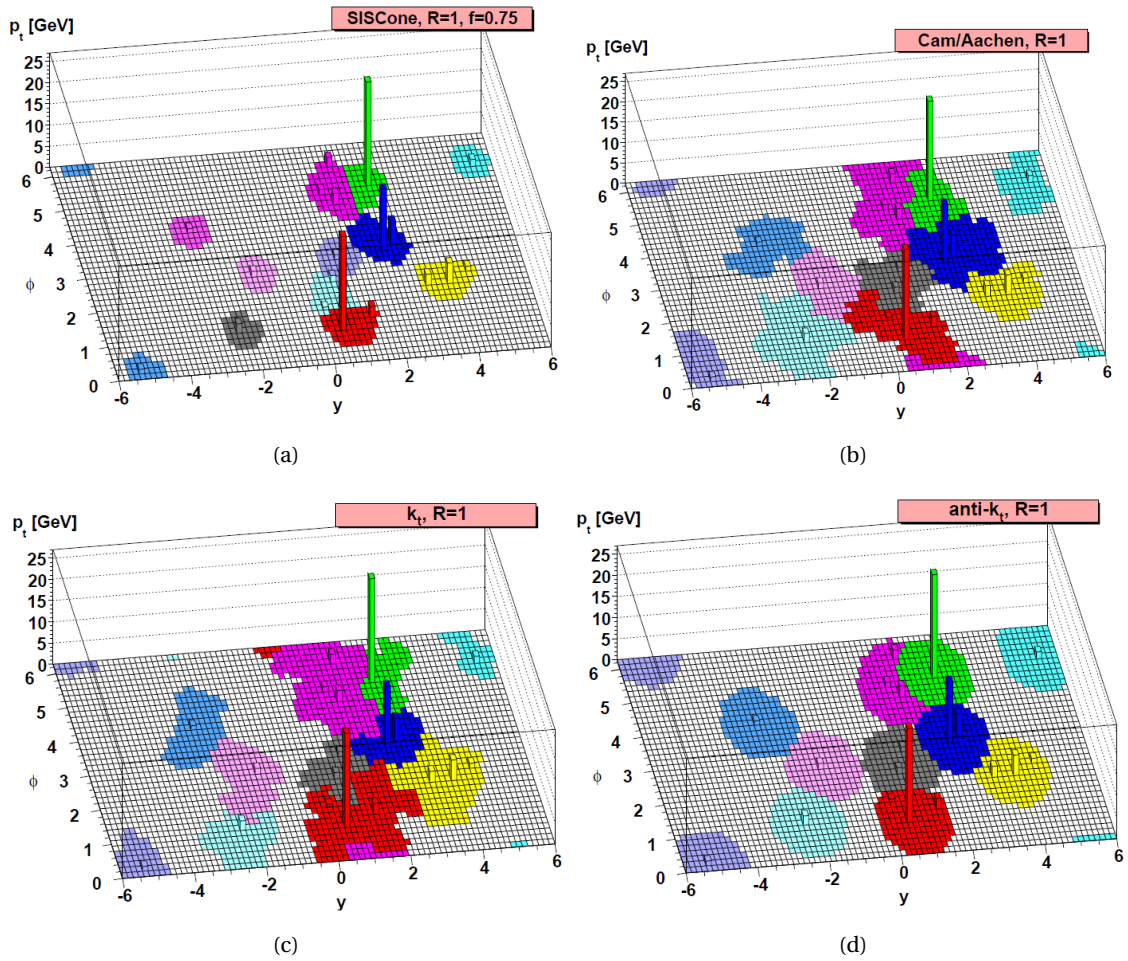
4.1.2 Topocluster Calibrations

EM Calibration

There are two possible levels of calibration we can apply to our topoclusters. The first of these is the EM calibration. The EM calibration works by recalculating both the signal and the average expected noise of each of the calorimeter cells to the response from electrons (see Section 2.7). This topocluster calibration aims, therefore, to reconstruct the energy deposits from electrons and photons more accurately. The EM calibration, however, does not aim to reconcile the effects of non-compensating calorimeter responses. It also does not take into account signal losses caused either by inactive material within the detector, or as a consequence of the topological clustering method.

Hadronic Calibration

The second calibration processes is the hadronic calibration. It does aim to attune the different calorimeter responses and the effects of signal loss, and therefore the corrections it produces are generally much higher than for the EM calibration. It includes cell signal-weighting, and is therefore known as the Local Cell Weighting (LCW) calibration. The hadronic calibration follows a number of steps: firstly the topoclusters are classified by their source, as either hadronic and EM. This is important as the calibration is different for each energy deposit. As mentioned in Section 2.2.4, the penetrating depth of a hadronic shower is much less than that of an EM shower. This mean that identifying a topoclusters as hadronic or EM in origin can be rooted in its longitudinal depth, λ_{clus} and the signal energy density within the cluster, ρ_{cell}. Both of these variables can be used to define a dynamic scaling, allowing the probability of a topocluster's origin to be determined, and therefore the most suitable calibration chosen.



(a) (b)

(c) (d)

Figure 4.2: Results of different jet clustering algorithms in MC simulated event [17]. Shown are the results of a cone algorithm, (a), the C/A algorithm, (b), the k_T algorithm, (c), and the Anti-k_T algorithm, (d). The beneficial circular nature of the Anti-k_T jets is visible here.

4.2 Jet Clustering

Now we have the inputs with which to reconstruct out jets, the next step in our jet reconstruction is to collate these topoclusters into meaningful structures. We must find way to decide which clusters are part of which jet. There are many possible ways to approach this problem and we require a solution that meets both theoretical and experimental requirements to building our jets. The combination of our jet algorithm and the parameters we choose for it forms the basis of our jet definition.

In 1990, the "Snowmass accord" [152] set out some general agreements on the properties of jet definitions. Some of the more important ones are as follows:

- Simple to implement within an experimental analysis;
- Simple to implement within theoretical calculations;
- Defined at any order of perturbation theory;

- Gives a finite cross section at any order of perturbation theory.

Our jet reconstruction process must also meet the requirements for Collinear Safety and be IRS (known as being Infrared Collinear safe (IRC)) [17]. For Collinear safety to me bet, the splitting of energy deposits from a large p_{T} particle into two separate collinear particles should not alter the results of our jet reconstruction. To be IRS soft particles added to or removed from the jet should not alter the results of our Jet Clustering algorithm. This prevents soft radiation from the pile-up or from the UE affecting the final state of our jet.

Beyond this, different jet algorithms can lead to numerous differences in the reconstruction results. Different jets can be produced from the same hard collision, including different topologies, a variety in the numbers of jets and differences in the shapes of jets. The choice we make for our algorithm is therefore important, and it is vital we take the objective of our analysis into consideration. We may select jet algorithms that give small area jets in the aim to reduce the effects of pile-up, or select larger jet radii for studies of boosted and heavy particles.

4.2.1 Cone Algorithms

Initial intuitive attempts at jet clustering lead to the development of cone-type algorithms, developed in the 1970s [153]. These jet clustering algorithms can be considered as "top-down": dependent on the concept that the process of hadronisation and QCD branching both leave the broad features of energy flow within an event intact. This allowed a fully consistent calculation in perturbative QCD as to the likelihood of multiple jets within an event. The process involves defining a cone about the collimated spray of particles within our event [154]. These algorithms are problematic however, as they do not consistently meet the requirements for Infrared Collinear safe (IRC) safety.

4.2.2 Sequential Recombination Algorithms

Pursuits for Jet clustering Algorithms that could overcome difficulties with IRC lead to the development of sequential Recombination algorithms. These have a "bottom-up" structure. The topoclusters in our calorimeter are combined in an iterative fashion, the order of which relies on a distance measure between pairs of topoclusters that can depend on a function of their p_{T}. This parameter relating to the momentum is used to determine jet resolution. Beyond jet finding they also define a sequence of clustering to our event, related to probabilistic models of parton branching, and necessary for JSS studies.

There are three such algorithms that are most often used in hadron collider experiments, sometimes known as the $\mathrm{k_T}$-like algorithms. Each of them rely on calculating the distance between pairs of clusters within a chosen parameter space to decide on which pairs should be recombined into 'proto jets'. The calculations required to follow these schemes can be summarised by three equations:

$$d_{ij} = \min\left(d_i, d_j\right) \frac{\Delta R_{ij}^2}{R^2} \tag{4.5}$$

$$d_i = p_{T,i}^{2p} \tag{4.6}$$

$$\Delta R_{ij}^2 = \left(\eta_i - \eta_j\right)^2 + \left(\phi_i - \phi_j\right)^2 \tag{4.7}$$

The first of our equations, Equation 4.5, sets the value of our distance parameter, d_{ij}, between two topoclusters within our calorimeter, i and j. It relies on R, a free parameter chosen by us to decide our desired jet size and R_{ij}. The final equation, Equation 4.7, describes R_{ij}: the distance in $\eta - \phi$ space between i and j. The second, Equation 4.6, describes the distance, within our parameter space, between the topoclusters, i and the beam of our collider. This is a function of the p_{T} of the particle and contains a value, p, which we are able to choose to determine our algorithm. The default value of p is one, and defines our first sequential recombination algorithm, the $\mathrm{k_T}$ algorithm.

The $\mathrm{k_T}$ Algorithm

The $\mathrm{k_T}$ algorithm [155, 156] combines together close and low-p_{T}, and therefore soft topoclusters, first. This is done by setting the value of p, describing the beamline-to-cluster distance within our parameter space found within Equation 4.6, to one:

$$d_i = p_{Ti}^2 \tag{4.8}$$

$$d_{ij} = \min\left(p_{Ti}^2, p_{Tj}^2\right) \frac{\Delta R_{ij}^2}{R^2} \tag{4.9}$$

Recalling Equations 4.5 and 4.6 with our given p-value allows us to calculate both the distance parameter between our two clusters, i and j, and between i and the beam. We do this for all clusters and find the minimum value. If the minimum in our list of distance values is a value of d_{ij} we combine the clusters i and j, forming our proto-jet. We then remove both of these clusters from our list, and add the new proto-jet to our list of clusters. If the minimum value in our list of distances is a value of d_i, our cluster i is labelled as a jet, and is removed from our list of clusters. After either of these decisions we compute all of our distances again and continue onwards until all of our particles are clustered into jets.

One of the consequences of this process is that all clusters will eventually become a jet, or part of a jet. Single soft particles can therefore also become jets, and this may be remedied by defining a minimum p_{T} threshold that our clustered jets must pass to be considered in further steps of our analysis. Unlike cone algorithms, the $\mathrm{k_T}$ algorithm is considered IRC safe, however it produces geometrically irregular jets, which can complicate certain reconstruction corrections. See Figure 4.2

The Cambridge/Aachen Algorithm

The Cambridge/Aachen (C/A) algorithm [157, 158] uses a distance measure that is only based on the angular distance between clusters. Depending only on geometric factors means it is

therefore independent of the energy and momentum of the particles being clustered. We can thus set the value of p in our distance measurements to zero:

$$d_i = p_{Ti}^0 = 1 \tag{4.10}$$

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2} \tag{4.11}$$

As all values of d_i are now equal to one, the sequence first sequentially combines together pairs of clusters that are closer that our maximum defined jet radius, R, and stops only when all jets are separated by a distance equal to, or larger than R. All remaining clusters are then defined as our final jets. The C/A also gives irregularly shaped jets, however the purely-geometric nature of the C/A algorithm makes it possible and easy to inspect a specific jet on many different angular scales – a property that will come in particularly handy in our analysis presented in Chapter 6.

The Anti-k_T Algorithm

As the k_T algorithm uses the distance parameter $p = 1$, the Anti-k_T algorithm [159], conversely, uses the distance parameter $p = -1$, such that:

$$d_i = p_{Ti}^{2p} = p_{Ti}^{-2} \tag{4.12}$$

$$d_{ij} = \min\left(p_{Ti}^{-2}, p_{Tj}^{-2}\right) \frac{\Delta R_{ij}^2}{R^2} \tag{4.13}$$

This leads to a sequential combination favouring high-p_T clusters combining with nearby clusters. This leads to jets growing outwards around a hard "seed", meaning a centre of high-p_T topo-clusters. The combined use of the energy and angle within the distance parameters means that the Anti-k_T algorithm leads to IRC growth of jets. Once all jets are separated by a distance greater or equal to R we have our final set of jets.

The outward growth of these jets also leads to them being circular in the y-ϕ plane, as seen in Figure 4.2(d), with a radius of size R. This is a reason to favour the Anti-k_T algorithm over the k_T or C/A algorithms, which both produce irregularly-shaped jets. Regular jet geometries lead to fewer complications in detector corrections, and corrections from non-perturbative origins. The containment of the total jet within the chosen boundary of R means that it is easier to define regions of the detector in which all jets are fully contained. Regularly shaped jets also make the process of pile-up suppression simpler: the area of the jet can be analytically calculated following merging or splitting. The lower memory requirements for the Anti-k_T algorithm also means that jet reconstruction can be faster and more efficient [160].

One downside to the method is that unlike the other k_T-like algorithms the Anti-k_T method of reclustering does not follow a physically meaningful recombination order: it does not follow the jet fragmentation through the separate reconstruction, and subsequent combining of soft subjets and hard subjets. For all the benefits above, however, and also the better trigger matching performance and greater stability under high levels of pile-up, the Anti-k_T method is chosen as the standard jet reconstruction algorithm to define jets within analyses at ATLAS.

4.3 Processing Jets

To ensure our jets meet the required standards for analysis there are multiple levels of processing we can apply to improve them before we use them. The processes we follow depends on the aims of the analysis we are undertaking. We must also take note that Jets are built in a range of sizes, also chosen to suit the analysis objectives. Different radii of jets are used to strengthen jet reconstruction in a variety of ways, the methods in which we process and treat these jets must therefore reflect this.

The ATLAS experiment makes the distinction between small jets and large jets based on radius. Small jets have a radius $0.2 \leq R \leq 0.4$, most commonly set to $R = 0.4$ and are constructed using the FastJet package [161]. Using small jets has the benefit of reducing the effect of pile-up in jet reconstruction when compared to large jets $R = 1.0$, which are commonly used for reconstruction of heavy particles. For the analyses in this thesis I have chosen to use small-jets for this reason.

4.3.1 Jet Calibration

Figure 4.3: Calibration stages for small jets.

The first step we take to process our jets is a series of calibrations to account for the effects of many procedures on our jets: detector response, pile-up, non-compensating calorimeter responses, differences between data and MC simulations and more. The calibrations applied to a jet are dependent on information, such as the jet radius of reconstructed jets. For the small reconstructed jets, of radius $R = 0.4$, the calibration process undertaken is the JES calibration [162]. The purpose of the JES calibration method is to correct the JES[163] of those jets reconstructed from clusters in the calorimeter, in order to match the energy scale of truth jets within MC simulations. The JES process consists of multiple phases, listed below.

Origin Correction

Calibrating our hadronic jets first requires taking our reconstructed jets, made from topoclusters, and making corrections for the origin. The purpose of this is to guarantee that our jet direction is correctly aligned with the primary vertex, such that it points towards the hard-scattering of our event and not simply the geometric centre-point of the detector (See Figure 2.4). This calibration can improve the spatial jet resolution in η and does not impact the energy of our jet. We can crosscheck this calibration by using jets reconstructed from clusters in MC simulations with truth jets, constructed from truth particles.

Pile-up Correction

The second stage in our jet calibration process aims to negate various consequences of pile-up [164]. When reconstructing our jets from topoclusters, energy originating from pile-up can be incorrectly included. The average energy contribution from pile-up is distributed in a uniform manner throughout the η-ϕ plane. The homogeneity of this contribution simplifies the process of subtracting additional p_{T} from reconstructed jets using Equation 4.14.

$$p_T^{correction} = p_T^{reco} - \left(\rho \times A\right) - \alpha\left(N_{PV} - 1\right) - \left(\beta \times \langle\mu\rangle\right) \tag{4.14}$$

Here ρ represents the median p_{T} density of the reconstructed jet. It is dependent on the NPV within our event, and is by association an indication to the extent of pile-up. A represents the area of our reconstructed jet, thus the first term to be subtracted from our p_{T} represents the combined pile-up contribution within our jet [165]. The second term uses the NPV and a constant, α, to calculate an event-specific pile-up energy contribution to subtract. The final term includes a second constant, β, and relies on our mean number of interactions per bunch crossing, $\langle\mu\rangle$, and relates to out-of-time pile-up within our event (see Section 3.1.5). Our second and third terms involve calculations of difference in p_{T} between the reconstructed jet and a truth jet.

MC-based energy calibration

The next step is an energy calibration, and uses information from our MC simulations, used to correct the four-momentum of our jet to match the energy scale of the particle-level jets [166]. We begin our correction by geometrically matching reconstructed jets to truth jets within the detector, within a distance of $\Delta R = 0.3$. With these jets matched to one another, we can use the difference in their energy to calibrate our reconstructed jets.

Another correction is possibly made at this stage, this time to the η direction of our reconstructed jet. Biases in η are present in some regions of the detector and are particularly prominent in jets passing through regions of the calorimetry system with variances in granularity, materials or geometry. Ultimately, differences in response in our reconstructed jet's energy can occur due to this trans-regional effect.

This step of the calibration is unique in that it is the only stage of our calibration process for small jets that is also used for large jets. Large jets are calibrated first through grooming (see section 4.6) then through MC-based processes to calibrate the energy of the jets. Large jets also undergo a MC -based mass calibration based on mass response, $R_m = \langle m_{reco}/m_{truth}\rangle$.

Global Sequential Calibration (GSC)

GSC is a procedure used to counter any content-dependencies that exists within jet reconstruction [167]. By decreasing the sensitivity of variances that exist between quark- and gluon-jets in calorimeter response, GSC aims to tackle dependence on the energy distribution and particle

composition in jet reconstruction. Jets originating from quarks most often possess hadrons that possess a significant portion of the jet p_{T} – these jets, therefore, highly penetrate the detector. The same is not true for gluon-jets, as these most often contain particles with a broader distribution of jet p_{T}. GSC therefore aims to improve resolution in JES.

Five different jet observables form the basis for this calibration process:

- The jet-energy fraction within the first layer of the HCAL;
- The jet-energy fraction within the final layer of the HCAL;
- The number of tracks with p_{T}>1 GeV associated to the jet;
- The track-width of the jet;
- The number of muon track segments associated to the jet;

The corrections are globally applied, for all regions of the detector, to the four-momenta of the reconstructed jet.

In-situ Energy Calibration

The final step in the JES calibration relies on residual in-situ energy corrections, accounting for disagreements between jet responses in data and MC simulations [168]. Reconstructed jets have corrections applied to their p_{T} through a process relying on a sequence of reference objects, each of which focuses on a different p_{T} range. Within the central region of the detector, $|\eta| < 0.8$, jets are calibrated using Z Boson (p_{T} of 20 GeV to 500 GeV), photon (p_{T} of 36 GeV to 950 GeV) and multijet (p_{T} of 300 GeV to 2000 GeV) events. Within the forward region of the detector, $0.8 < |\eta| < 4.5$, jets are calibrated using well-measured jets from the central region using dijet events.

4.3.2 Jet Cleaning

Identifying jets with origins external to our collision event, or from noise within the detector, relies on given requirements on quality. Events possessing more “bad” (fake) jets than good are removed from analyses [169]. The main sources of these fake jets include the background from upstream proton losses induced by the beam in the LHC, cosmic rays and noise within the calorimeter. Jet cleaning makes use of multiple jet-quality variables, in order to identify good and bad jets. Bad jets originating from noise in the LAr can be identified by their signal pulse shape within the calorimeter. Jets from any of the external sources can be recognised through use of tracking info and energy ratios. Both loose and tight selections exist for distinguishing good jets from bad. The loose working point, which is most commonly used in physics analysis, has an efficiency above 95 %.

4.3.3 Rejecting pile-up Jets

The jet reconstruction process necessitates the effective rejection of jets originating from pile-up [170, 171]. These pile-up jets are partly removed through the calibration process, as this process reduces the p_{T} of these jets so that they are below the p_{T} threshold. Jets that remain can be removed through the use of the Jet vertex Tagger (JVT) discriminant following JES. The JVT is a multivariate measurement constructed from observables relating to the fraction of charged tracks that are both within the jet and point to the PV – an indication of their originating from the hard-scatter within our event. This process, carried out in both MC simulation jets and jets from data, makes use of information from jet tracks originating from the hard scatter, and relates to the fraction of jet p_{T} they carry. This algorithm targets low-p_{T} jets (p_{T}<60 GeV). These jets must be central within the detector, $|\eta| < 2.4$, and be matched to tracks passing a p_{T} threshold of 0.4 GeV.

We can also use the number of tracks associated to a jet to identify its origin, such as QCD jets from vertices caused by pile-up. The ID only reconstructs tracks that are in-time with the event, and these jets will have fewer tracks associated to them. Pile-up jets originating from local fluctuations, known as Stochastic Jets, are superpositions of jets, and can be identified through the multiple vertices their tracks originate from. Many vertices, leading to a JVT score below the defined threshold, indicate a Stochastic Jet, which is subsequently removed.

4.4 Flavour Tagging

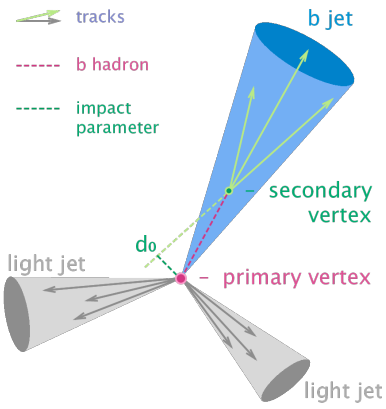
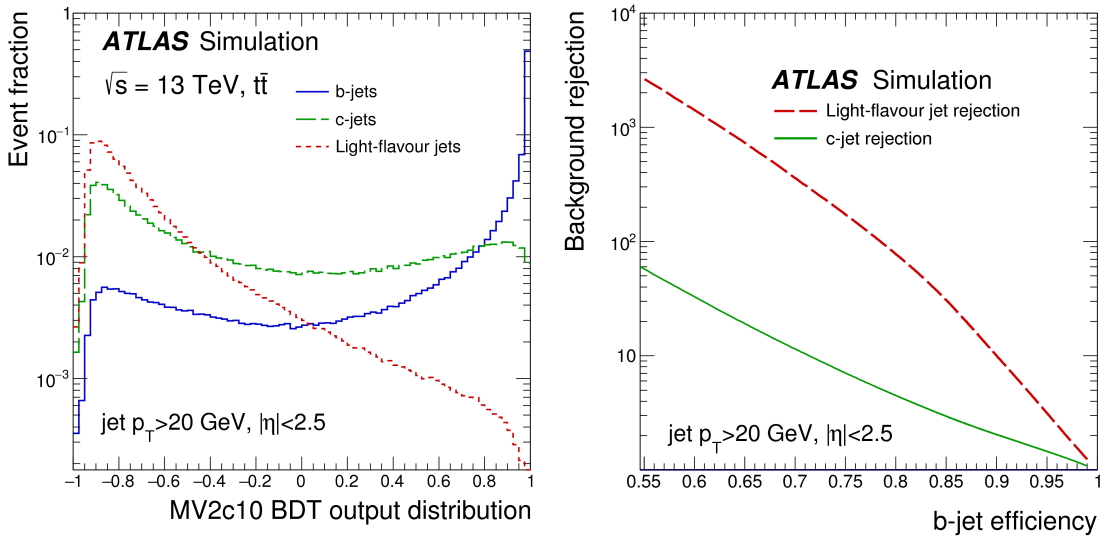


Figure 4.4: Illustration of the key criteria utilised in identification of jets initiated by b-quark decays. d_0 is the inverse IP [18].

Identifying the source of our jets is a crucial aspect of physics analysis, allowing us to identify rare (or even undiscovered) particles and interaction processes. The procedures for classifying our jets by their particle of origin are known as tagging. A particularly prominent area of flavour-tagging for small jets is b-tagging, used to identify jets containing b-hadrons. B-hadrons have relatively long lifetimes for hadrons (≈ 1 ps), therefore their prolonged journey through the detector leaves distinctive secondary vertices with high- IP tracks (see Figure 4.4). We define the distance of closest approach between a chosen track and its associated primary



(a) (b)

Figure 4.5: From [19], MV2c10 BDT score for b-jets (solid line), c-jets (dashed line) and light-flavour jets (dotted line) in simulated $t\bar{t}$ events, (a), and the light-flavour jet (dashed line) and c-jet rejection factors (solid line) as a function of the b-jet tagging efficiency of the MV2c10 b-tagging algorithm, (b). Rejection, here, is defined as the inverse of the efficiency tagging of a c-jets (dashed line) and light-flavour jets as a b-jets.

vertex as the IP. These high- IP tracks can be matched to jets, permitting selection criteria to identify and distinguish these "b-jets" from other flavours. The performance of b-tagging at ATLAS is therefore intimately linked to and reliant on tracking ability in the ID particularly the IBL (see Section 2.2.3).

Beyond tracking, algorithms are also employed to classify and flavour-tag b-jets. The ATLAS collaboration employs three distinctive categories of algorithm for b-tagging purposed, each attempting to recognise b-jets though different methods.

- **IP2D and IP3D**: These rely on the IP of tracks matched to the jet, as IP should be larger for tracks originating from the decay of b-mesons. The probability distribution of two variables, the significance of the IP and longitudinal significance are used to generate a discriminant. IP2D utilises only the impact parameter significance, d_0/σ_{d_0}, for a discriminating variable, whereas IP3D also makes use of the longitudinal impact parameter significance, such that it's discriminant is $z_0 \sin\theta / \sigma_{z_0 \sin\theta}$ [172].

- **Secondary Vertex Finding Algorithm**: This method reconstructs secondary vertices within the cone of the jet. During reconstruction pairs of tracks forming two-track vertices are rejected if they are compatible with a background source, such as photon conversions, decay of lon-lived particles or hadronic interactions with the detector [173]. Other criteria aim to prevent fake vertex reconstruction. Much higher rates of reconstructed secondary vertices distinguish b-jets from both c- and light-jets.

- **Decay Chain Multi-Vertex Algorithm (JetFitter)** [174]: This reconstructs the full decay-

chain of b-hadron within the jet based on the jet's topological structure. The trajectory from the primary vertex, through to the b-hadron, and it's subsequent decay products are extracted. This is undertaken through the assumption the primary, secondary and tertiary vertices of the hard-scattering, b-hadron and c-hadron respectively, line up, and an approximation of the flight direction of the b-hadron. The vertices and tracks of the b-hadron and c-hadron are therefore identifiable.

Additionally to these methods there exists the multivariate method of b-tagging [175, 176, 177]. This multivariate algorithm combines the three previous tagging methods using a BDT, producing a multivariate tagging value with the best separating power. This algorithm comes in three forms: MV2c00, MV2c10 and MV2c20. Each of these are named after the percentage of c-jets used within the training samples in addition to the majority of light-flavoured jets. MV2c10 is trained on $t\bar{t}$ events that contain at least one lepton. 7 % of jets within these samples are c-jets, and therefore have c-hadrons.

These c-jets are a source of background in our searches for b-jets (see Figure 4.5). Although the lifetime of a c-hadron is shorter, and therefore the distance traversed within our detector is shorter, they leave similar signals, including the indicative secondary vertex. Use of the MV2c10 algorithm is therefore critical to discriminate between b-jets and c-jets, as well as jets from light quarks (u, d, s), jets from gluons, or jets originating from hadronically decaying τ -leptons.

The MV2c10 algorithm is calibrated with four different fixed WPs, each of these have a given b-tagging efficiency and rejection factor for c-, τ -, and light-jets. Efficiency in tagging at a given working point is $\epsilon_b = N_b^{tagged}/N_b^{true}$. The rejection factor of a WP is taken as the inverse of the tagging rate of background objects, for example the rejection factor of c-jets is given as $R_c = N_c^{true}/N_c^{tagged}$. These WP are defined by a single cut value on score sourced from a BDT using 21 different jet variables. These 21 consist of variables from a combination of the three different, previously defined b-tagging algorithms.

WP	BDT cut	b-jet efficiency	c-jet rejection	light-jet rejection	τ-jet rejection
Very tight	0.9349	60 %	34	1538	184
Tight	0.8244	70 %	12	381	55
Medium	0.6459	77 %	6	134	22
Loose	0.1758	85 %	3.1	33	8.2

Table 4.1: WPs for MV2c10 b-tagging algorithm, shown with respective b-jet efficiency and other jet rejection rates attained from training on $t\bar{t}$ events.

Another tagging technique with growing relevance is the use of jet substructure (JSS) observables to increasingly improve the capabilities of flavour tagging [178, 179].

4.5 Jet Substructure

Following, or during, the process of Jet Reconstruction, we can start to look in closer detail at information within our jet, giving us clues to the physics of our event. Understanding the flow and distribution of particles within our jet to identify its origin relies on examination of the jet substructure (JSS). How beneficial this JSS is to deepening our understanding of the physics relies on a number of factors, such as how able we are to resolve decay structures within our jet, or our ability to distinguish and remove contributions from pile-up. Complexities in accurately determining JSS stem from limitations in detector resolution (either in space or in energy), deficient calorimeter acceptance of low energy particles and extra contributions from pile-up. We can undertake this task by analysing JSS observables related to:

- The number of identifiable constituents within the jet;
- The energy of these constituents;
- The angular separation between these constituents in our detector geometry.

As an example, the identification of the number and nature of decay prongs - signalled by smaller, hard (high-p_{T}) jet-like structure within jets - can assist us in identifying the origin of the jet. Jets with little, or no sign of this hard-substructure most often originate from decays of gluons or light-quarks. In ATLAS, this can be measured by using the "N-subjettiness" algorithm [179].

4.5.1 Jet Mass

One of the more valuable JSS observables, and of particular focus for this thesis, is jet mass. Outside of heavy quarks, partons can generally be considered to be massless, however jets are not massless – especially any jets with notable levels of JSS. Analysing the invariant mass of our jets and comparing it to objects we aim to tag can help us discover the source of hadronisation and boosted hadronic decay.

In a perfect reconstruction of the physics of our event the jet mass that is produced by recombining the four-momenta of its constituent should reflect the original source of the jet. We therefore sum the four-momenta of the jet constituents in the following way

$$m_{jet} = \sqrt{\sum_{i \in J} E_i^{\,2} - \sum_{i \in J} \overrightarrow{p_{\mathrm{T}}}_i^{\,2}} \tag{4.15}$$

Here J is our jet, constructed from constituents, i. Each constituent i has energy E_i and transverse momentum $\overrightarrow{p_{\mathrm{T}}}_i$. These constituents can be massive (such as tracks or massive particles) or can be massless (such as our calorimeter clusters or massless particles), and can be reconstructed into jets using any of the standard definitions. We can calculate the jet mass both before and after Jet Grooming (See Section 4.6) takes place.

Jet mass is most significant for those jets originating from the decays of heavier particles. As a rule of thumb, the opening angle of a dedcay can be given by Equation 4.16.

$$R \approx \frac{2m}{p_T} \tag{4.16}$$

For jets originating from light quarks or gluons it is less well defined, but scales proportionally to the p_{T}.

For the best results, we desire that the internal structure of jets remain intact throughout reconstruction, we therefore require a high level of resolution of the energy flow within jets. The jet mass is therefore vulnerable to detector effects, such as lateral energy spread. Single particle cascades can also negatively affect the resolution of jet mass. Accurately identifying jet mass is also complicated by the presence of particles that radiate out of, and then radiate back into, a jet. In theory these are represented by resummation terms called Non-Global Logarithms (NGL).

Further JSS analysis can help with adressing the effects if pile-up, improve measurement of mass and analysis of prong structure. We can also mitigate these issues through the use of Jet Grooming.

4.6 Jet Grooming

Reconstruction of hadronic jets from calorimeter clusters will characteristically necessitate grooming to counter non-perturbative effects degrading jet resolution. This source of these issues includes contributions from pile-up and other soft-radiation originating from sources other than the hard scattering in our event. The jet grooming process aims to leave only the intrinsic structure of the jet behind, and remove all other superfluous information. The various grooming processes undertaken can also distill the discrete nature of energy deposits within jets, allowing improved study of JSS variables for flavour assignment. A number of techniques are employed to groom jets, each with strengths and weaknesses.

4.6.1 Jet Filtering

Jet filtering [180, 181] requires we start with Jets reconstructed with angular size R_j. We then recluster the contents of our jet on a smaller angular scale, $R_{filtered} < R_j$. From here we choose the $n(= 1,2,3\ldots)$ hardest subjets, and reject all other clusters within our original jet. This allows selection of the dominant radiation products from a chosen decay, whilst removing much of the contamination from the UE.

4.6.2 Mass Drop

For the Mass Drop method of Jet Grooming we begin by reclustering a jet using C/A algorithm [180, 181]. Following this we reverse the last step of the clustering, leaving us with two sub-jets. We subject this pair of sub-jets to a requirement test relating to their mass, p_{T} and respective

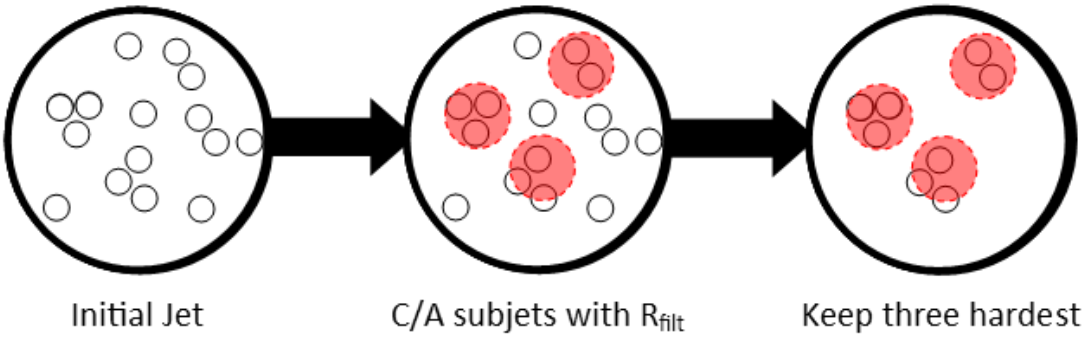


Figure 4.6: Jet "Filtering" process.

angular distance. We first ensure that the heavier of the two sub-jets has a significant "Mass Drop" from that of our ungroomed jet, μ_{frac}, shown in Equation 4.17.

$$\frac{max(m_1, m_2)}{m_{jet}} < \mu_{frac} \tag{4.17}$$

Here μ_{frac} is a chosen fraction of the original jet mass (e. g. 2/3). We next ensure that the splitting of the p_{T} distribution between our two sub-jets is not too asymmetric:

$$y = \frac{\min\left[{p_{T,1}}^2, {p_{T,2}}^2\right]}{m_{jet}^2} \times \Delta R_{1,2} > y_{cut} \tag{4.18}$$

y_{cut} is our chosen threshold (e. g. 0.09). If our sub-jet pair fails to meet these two requirements we discard the least massive, redefine the most massive as our new jet to be groomed, and we repeat the process. When we find a sub-jet pair passing our requirements we take the clustered pair to be our new groomed jet. In previous Higgs studies in ATLAS, a combination of Mass Drop and Filtering has been applied [181]. In this instance, the successful application of the Mass Drop grooming is followed with Filtering, using the distance between the two sub-jets, R_{12}, to define a new angular distance, $R_{filter} < R_{12}$, with which to recluster the new jets. As with jet filtering, the desired n-hardest remaining sub-jets are chosen, and all other clusters rejected. This allows selection of the dominant radiation products from a chosen decay, whilst removing much of the contamination from the UE. Here $R_{filter} = \min(0.3, \Delta R_{12})$ and $n = 3$ in order to reconstruct two b-jets and a gluon in ($H \to b\bar{b}$) searches.

4.6.3 Jet Trimming

The jet grooming procedure most commonly used within ATLAS is Jet Trimming [180, 182]. The method of jet trimming starts with reclustering the constituents a large jet (with $R = 1.0$) into subjets of chosen size R_{sub} using the kT jet clustering algorithm. To improve the rejection of soft radiation within our jet grooming, a suitable subjet parameter threshold as a fraction of our total jet p_{T} is chosen named f_{cut}. All subjets that fail to meet this p_{T} requirement are rejected:

$$p_T^{subjet} > f_{cut} \times p_T^{jet} \tag{4.19}$$

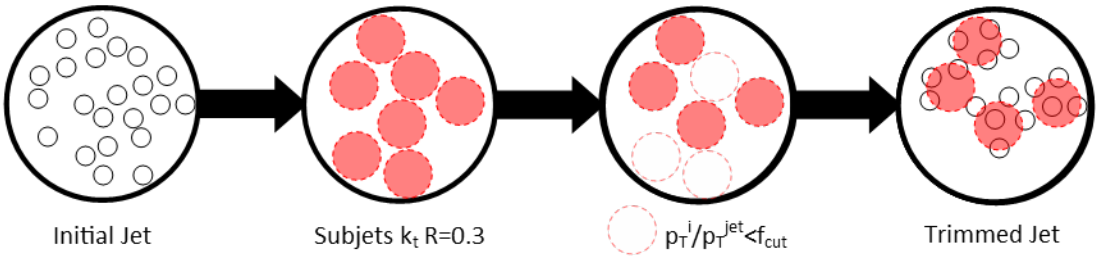


Figure 4.7: Jet "Trimming" process.

The jet is then reclustered using only those subjets that remain. Typical bounds used within ATLAS are $f_{cut} = 5\,\%$ and $R_{sub} = 0.2$ [180, 182]. This jet trimming algorithm aims to remove soft radiation and contributions from pile-up, the UE, ISR and MPI.

4.6.4 Jet Pruning

Another Jet Grooming technique is Pruning [180]. Here the process starts by reclustering the original jet using either the C/A or k_T algorithm. If the two sub-jets we are about to cluster together have an angular distance $\Delta R_{12} > R_{prune}$, and the softer of the two subjets has a p_{T} below our defined z_{cut} threshold

$$\frac{\min\left(p_{t1}, p_{t2}\right)}{p_{t1} + p_{t2}} < z_{cut} \tag{4.20}$$

then the softer of the two subjets is removed, and the next step of reclustering is moved onto. Pruning aims to enhance the substructure resolution within the jet by removing soft, wide-angle radiation at each clustering step.

4.6.5 Limitations With Grooming

Issues and limitations with current grooming methods have lead to searches for new ways to groom jets. The introduction of new, tuneable degrees of freedom introduces extra complexity that requires further investigation for analysis-specific optimisation. Jet grooming methods also require consideration of the presence and effects NGL – a problem specifically tackled by the use of soft drop grooming (soft drop grooming).

4.7 SoftDrop Grooming

Total and accurate predictions for mass, and other JSS variables beyond Leading Logarithm (LL) has so far not been possible, due to the presence of NGL, resummation terms associated with particles that radiate out of, and back into, a jet [183]. These additional radiative terms can preclude precision calculations and measurement of JSS variables. The development of novel Jet Grooming techniques, however, can allow us to solve these problems. Soft drop grooming is

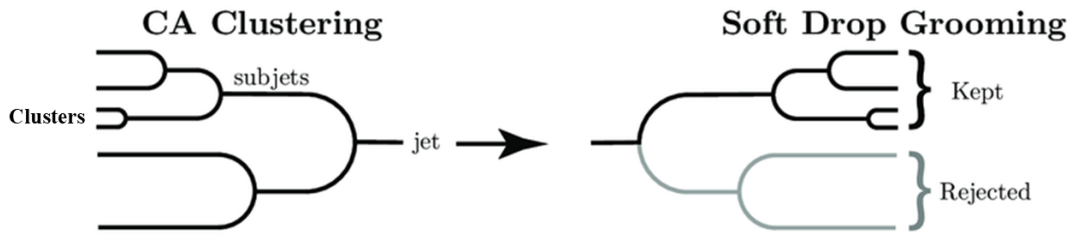


Figure 4.8: Schematic of both C/A clustering and soft drop grooming algorithms. The softest subjet of any pair not passing soft drop criteria is rejected. Image altered from [20]

a jet grooming procedure to remove soft and wide-angle radiation from the jet that is formally insensitive to NGL, allowing comparison to theoretical calculations at higher logarithmic orders [23, 20]. The calculation of jet mass from jets with soft drop grooming applied is, therefore, unaffected by NGL, opening up new ways to determining precision JSS.

Soft drop grooming requires the reclustering of Anti-k_T clustered jets using the Cambridge/Aachen (C/A) algorithm, clustering pairwise by distance starting with the closest pair of clusters within our defined jet radius. This newly reclustered jet is then declustered into the last pair of sub-jets to be clustered together. The pair of sub-jets remaining are then subjected to a test, to see if together they pass the soft drop condition, defined in Equation 4.21.

$$\frac{min(p_{T,j1}, p_{T,j2})}{p_{T,j1} + p_{T,j2}} > z_{cut} \frac{\Delta R_{12}}{R}^{\beta} \tag{4.21}$$

This means that, from our pair of sub-jets, that with the lowest value of p_T must have a calculated portion of the total sum p_T of the pair. Here the value z_{cut} is the scale of the energy to be removed through soft drop grooming, and β is the chosen sensitivity to wide-angle radiation. ΔR_{12} is the distance in $\eta - \phi$ space of our two sub-jets, and R is the radius of our reconstructed jet. If the p_T of the softer sub-jet is less than the required portion then we remove it from our jet, and the remaining sub-jet of the pair becomes our new, groomed jet. This process is applied again upon the new, groomed jet, and so forth, until either all parts of the jet are groomed away, or we reach a sub-jet pair that pass the requirements, whereupon we stop grooming.

THE ATLAS TRIGGER SYSTEM

5

One essential object is to choose that arrangement which shall tend to reduce to a minimum the time necessary for completing the calculation.

Ada Lovelace

The rate of data delivered by the LHC to ATLAS surpasses available recording and storing capabilities [12, 184]. Physics processes of interest are only produced in a minor fraction of these events, as most bunch crossings yield only soft hadronic collisions of no interest to physics analysis. A trigger system, therefore, is required to make rapid decisions based on event information. This ATLAS trigger system is utilised to select events of interest to keep for physics analysis, and discard others. This consists of fast real-time decisions based on signals from the various sub-detectors. The ATLAS trigger is designed to reduce the event rate from the 40 MHz bunch crossings from the LHC down to 1 kHz, a rate that is manageable for recording. Only data from events identified as containing potentially interesting physics, and of high energy are stored.

A strong understanding of how to best trigger on events of intersest is, therefore, cruicial to the efficient and effective capture of data from the detector. An increasin in luminocity of the ATLAS experiment, with a higher rate of proton-proton collisions, is likely to make this process increasingly complex [185] [186].

The first section of this chapter discusses the structure of the ATLAS trigger system and the second section covers analysis work done with the Jet Signature Trigger group to better understand how the trigger uses track finding for jet reconstruction.

5.1 The Architecture of the ATLAS Trigger System

An overview of the ATLAS Trigger and Data Acquisition (TDAQ) system is displayed in Figure 5.1. The system consists of two levels – the first uses custom fast electronics hardware to gather

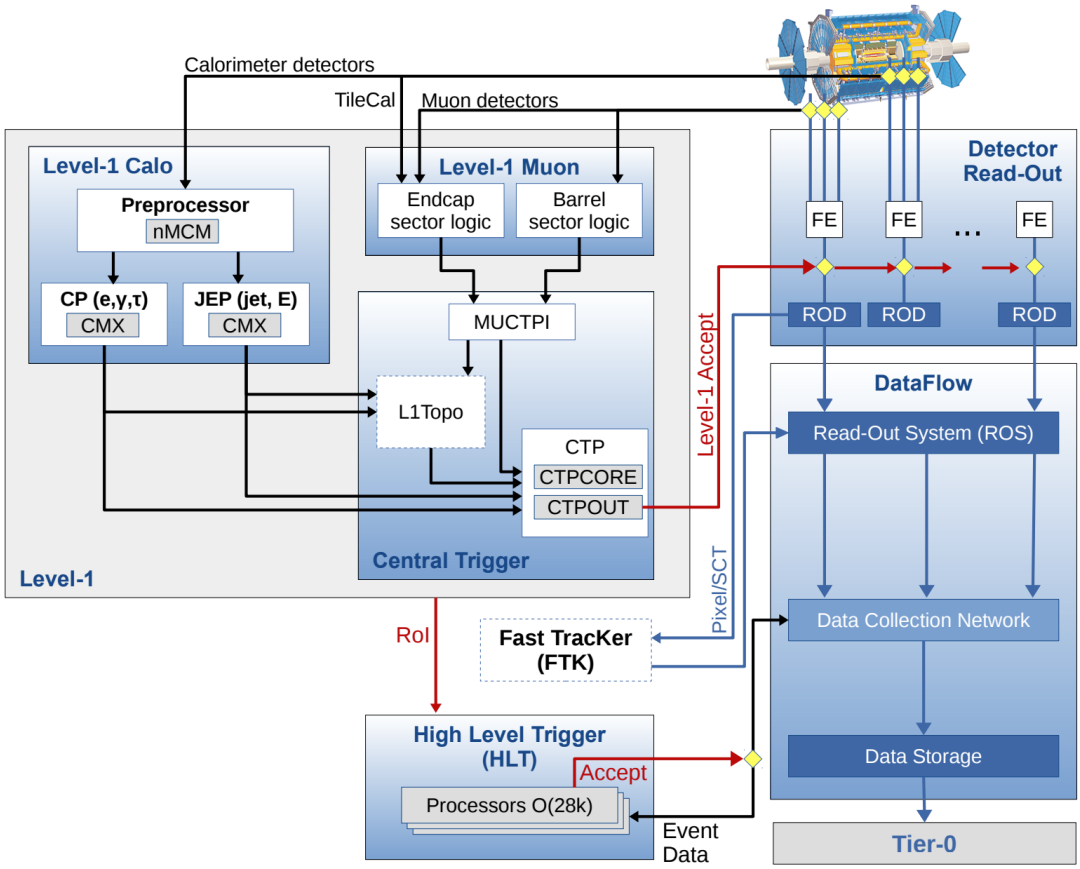


Figure 5.1: The schematic of the ATLAS TDAQ system [21].

coarse data at high speed, and the second uses slower, more complex software decisions and a finer granularity. The first, hardware based stage is the Level-1 (L1) trigger [187]. It uses information from the calorimeters and MS and identifies Region of Interests (RoIs). Event data taken from other sub-systems are stored in memory temporarily whilst L1 comes to a decision.

The second is called the High Level Trigger (HLT) [188, 189]. This is software based and runs algorithms using data from the whole detector, bringing in information from the ID and reconstructs events of interest. The HLT consists of software algorithms running on a cluster of computers on several thousand Central Processing Unit (CPU)s (HLT farm). During Run 1 the HLT was split into two parts, Level-2 (L2) and the Event Filter (EF), however to reduce complexity of the system, and increase efficiency in algorithms, during Run 2 these were merged into a single HLT stage [190, 191].

Each stage of Trigger system builds on the decision from the previous level, reducing the rate of data recording. Any objects that do not meet the requirements of the trigger are permanently discarded, therefore it is vital to select as many events with potentially interesting physics as possible in an unbiased fashion. The triggers identify signatures of potentially interesting physics, such as high p_{T} particles, jets or large $E_{\mathrm{T}}^{\mathrm{miss}}$. These are used for benchmarks for event selection.

There exists a trigger menu, comprised of a list of trigger selections to be active at certain

times during the data-taking period. The menus consist of the list of L1 and HLT triggers to be running during a particular luminosity block of the LHC. Trigger items can have a pre-scaling. A pre-scale of 5 means that only 1/5 events passing the selected trigger is kept. This is done randomly and allows for an optimal usage of the bandwidth whilst luminosity within the LHC changes. If the output rate is too high, a determined chain of decisions can be pre-scaled by a factor P: only 1/P events that meet criteria are accepted.

There is also the Fast TracKer (FTK) system, which was installed in Run 2, used to process events accepted by L1, and to seed HLT algorithms [192, 193]. It provides global track reconstruction from within the ID at L1 trigger rate using lookup tables stored in associative memory chips custom made for pattern recognition.

All accepted data are eventually passed through the TDAQ system into data streams for offline physics analysis, trigger level analysis and monitoring or used for detector calibration. This channels data from sub-detector to permanent storage, where an offline computing infrastructure of four tiers (Tier-0,1 2 and 3) is used to construct the data from streams into meaningful event information. The data flow through these tiers is managed by the Data Acquisition system.

5.1.1 Level 1 Trigger

The Level-1 (L1) Trigger uses custom-made fast electronics to complete the first stage of event selection based on information from the sub-detectors [187]. Its decision is then passed onwards to the Central Trigger Processor (CTP). L1 Trigger is hardware based and is required to make decisions within 2.5 µs to reduce data rate down from 40 MHz to a maximum of 100 kHz.

To do this, the L1 trigger defines one or more RoIs in the η-ϕ plane containing signatures of interesting physics. RoIs incorporates information about the type of object detected, its energy and its coordinates within the sub-detector and are used later for reconstruction and tracking within the HLT. RoIs are defined by the two sections of the L1 trigger: L1 Calorimeter (L1Calo) and the L1 Muon (L1Muon). Construction of these RoIs rely on the calorimeters L1 Calorimeter (L1Calo) and the MS, L1 Muon (L1Muon) triggers. These are used to decide if particles within the detector originate from high-energy collisions containing interesting physics. The L1Calo function of the trigger searches the calorimeter for signals to identify large Transverse Energy (E_{T}), $E_{\mathrm{T}}^{\mathrm{miss}}$ and high-p_{T} objects. It completes this either by using low granularity trigger towers (measuring energies of electrons, photons and jets) or by ascertaining weak interactions have taken place (recognising hadronically decaying tau leptons using their showers. Muons are detected in the independent trigger system within the MS through the L1Muon trigger. This also determines electroweak interactions have taken place when triggering on muons within the the MS.

Since Run 2 there exists new Level-1 Topological (L1Topo) triggers [194]. These were introduced with the objective of reducing the event rate before the Central Trigger Processor (CTP). This is particularly important at increasingly high luminosities. When using only the L1Calo and

L1Muon triggers, the CTP can only base decisions on information relating to the multiplicity and thresholds of potential physics objects. The new L1Topo trigger can improve upon this by accessing higher-level topological information including invariant mass and angular variables. It combines kinematic information from the L1Calo and L1Muon, allowing more complex selections based on event topology. The L1Topo processes data through algorithmic firmware. These are loaded onto integrated circuits known as Field Programmable Gate Array (FPGA)s before being sent onwards to the CTP. L1Muon uses firmware to send coarse granularity measurements of η-ϕ and p_{T} onwards to the L1Topo.

As information from all L1 sections reach the CTP, the trigger menu is implemented. This consists of 512 distinct items, each of which is a logical combination of criteria to be achieved for an event to be accepted. This can be based on thresholds, multiplicities and L1Topo flags. When an object in the sub-detectors meets these requirements, L1 trigger sends an "accept signal". The event information is then buffered into the Read-Out System (ROS) and sent alongside any RoI data to the second level of the ATLAS Trigger: the HLT. At this point data from the full detector is read out. This is processed by fast electronics and the information is stored in front-end pipelines on or near to the detector. This is stored here until further processing is required of the data. The data is sent to Read Out Driver (ROD)s, specific for each detector, and is later passed onto the Read-Out System (ROS).

The L1 latency is measured from the time of the p-p collision, during the bunch crossing, up until the L1 trigger decision is ready to be sent to the front-end electronics. The L1 latency is required to be less than 25 µs to match the pace of these collisions. There exists a number of procedures in place to slow the flow of data to prevent the overlapping of readout windows: the CTP prescribes two different types of "dead time" within the processing; simple and complex. Simple dead time sets a limit on the minimum time between two consecutive accept signals from the L1. Previously this was the time span of 5 bunch crossings, but is now the time span of 4, 100 ns. Complex dead time sets a maximum on the number of L1 accepts within a given period of time. Any signals during this time period are rejected by the CTP. The number of the luminosity block within which the data is taken is kept.

5.1.2 High Level Trigger

The second part of the ATLAS trigger system is the High Level Trigger (HLT) [188, 189]. The HLT is a software-based system and uses RoIs selected by the L1 trigger at full granularity and precision. The HLT implements additional selection criteria to data coming from the L1 trigger and the rate is reduced from 100 kHz to 1 kHz in ≈250 ms. This stage runs on several thousand CPUs. Following acceptance by the HLT trigger further algorithms are implemented the data for online analysis, fully reconstructing events to wholly recreate the physics objects within.

During Run 1, the HLT consisted of two separate stages; L2 and the EF. Before Run 2 these were combined into a single farm to reduce complexity, remove redundancies between them and allow for dynamic sharing of resources between algorithms [190, 191]. The inclusion of

tracking information from the ID at the HLT stage of triggering enables identification of objects such as electrons and muons. Tracking within the HLT has two subdivisions: fast tracking and precision tracking. Fast tracking allows the rejection of the majority of events, and is utilised to reduce the overall processing time. Precision tracking goes on to reconstruct the events that remain. During data taking through 2015 and 2016 the software based Fast Track Finder (FTF) performed fast tracking. The Fast Track Finder (FTF) delivered approximately 2,500 independent trigger chains. Each of these trigger chains are a sequence of algorithms for pattern recognition. These algorithms are executed within the RoIs provided by the L1 trigger and were specially developed both to provide precision reconstruction of tracks from well-separated objects, such as electrons photons, muons and taus, and to match these tracks with finer granularity information from the calorimeters.

Muon candidates, following reconstruction accomplished solely from information coming from the MS, are linked with RoIs chosen from the FTF. Information from the MS was used within the FTF to identify muon candidates only indicated within the MS. The FTF would then use this to identify ID tracks within the provided RoI.

Once an event has been accepted by the HLT it is transferred to local storage and is then transmitted to the computing centre of CERN where offline reconstruction takes place. This offline analysis employs complex algorithms over full events to reconstruct candidate physics objects. These events are written to different data streams, each of which will be used for physic analyses, trigger level analysis and monitoring or for calibration of the detector. These streams consist of a single jet/tau/$E_{\mathrm{T}}^{\mathrm{miss}}$ stream, a muon stream, an electron/photon stream and a minimum bias stream. Events can feature in more than one stream: these streams are therefore defined as inclusive. Overlap is kept to a minimum, with only about 10 to 15% of events duplicated. Only for physics analysis is the full event information written to the stream. This allows a lower bandwidth as only necessary information is written for other analyses.

The Jet Signature Trigger

The LHC is a hadron collider, and as such produces a very large number of hadronic jets, in fact most events collected by ATLAS contain jets [195]. The ATLAS HLT *jet signature trigger* uses high-speed jet reconstruction algorithms to trigger on events with chosen hadronic decay signatures. Developmental work on the ATLAS HLT jet trigger focuses on improving the efficiency and accuracy of these methods to better identify and trigger on events of most potential interest to hadronic analysis.

The jet trigger selects for both high-p_{T} jets, for searches involving high-mass exotic states, and also selects for lower-p_{T} jets for use in calorimeter calibration and of offline jet calibration. The jet trigger selects chosen events through detection of jets reconstructed and passing selection cuts. Cuts on the p_{T} and η range are of most importance within the jet trigger, making accurate reconstruction of the jet energy and direction critical. As jets are defined by the reconstruction algorithm used to produce them, comparison with offline jet reconstruction used in

analyses is essential to evaluate the jet trigger.

The ATLAS Fast TracKer

In 2017 the installation of a new, high-speed hardware-based tracking system, the Fast TracKer (FTK), took place [192, 193]. The motivation for the Fast TracKer (FTK) was to improve tracking and tackle a number of developing challenges, both by reducing the computational load of track-finding, likely to be exacerbated by future increases in LHC luminosities [185, 186], and by improving tracking precision through the use of global, rather than local, track reconstruction. Rapid pattern recognition, relying on particle look-up tables found in hardware memory chips, and reconstruction capabilities with rates approaching those of the L1 trigger gives the High Level Trigger (HLT) early access to global tracking information.

The precision tracking software is based on algorithms using track information, rather than just pattern recognition. Access to full object reconstruction, seeded by spacial coordinates and refined tracks from the FTF, prevents the precision tracking from being limited by RoIs. Access to the full range of information from the calorimeters enables establishment of global quantities, such as $E_{\mathrm{T}}^{\mathrm{miss}}$ and full reconstruction of jets within the HLT Jet Signature Trigger. Detection of b-quark decay, through recognition of secondary decay vertices, is vastly improved through this precision tracking.

5.2 Fast Tracking in the ATLAS HLT Jet Trigger

This section gives an overview of the technical work performed to obtain the ATLAS authorship qualification. The work was done within the Jet Signature Trigger group detailed in Section 5.1.2, and covered use of the FTK.

Increases in luminosity in the ATLAS experiment leads to some detriment in jet reconstruction: the increase in proton-proton impacts lead to higher likelihoods of misidentifying information from collisions in addition to the collision of interest, also known as pile-up. As proposals for future increases in luminosity in the LHC are likely to exacerbate this problem, solutions must be found [185] [186]. One process likely to improve discrimination, therefore reducing the effects of pile-up, is the increased use of identifying particle tracks for jets. Although, due to long processing times, this has previously been limited to b -quark tagging for b-jets built around RoI in the detector (See Section 4.4).

The ATLAS Fast TracKer (FTK) (Section 5.1.2) is a system that does global track reconstruction after each L1 trigger, to enable the HLT to have early access to tracking information [192]. It allows the possibility of rapid b-quark and τ-lepton identification within the online trigger system by providing a list of tracks with excellent resolution at the start of HLT processing. These improved, and now global, track reconstructions allow more information to be assigned to a jet. Investigations into FTK vertices, jets and tracks, and how they vary from those constructed both within the HLT and within offline reconstructions can give us more insight to the changes

this improved tracking brings. Thus, I begin an exploration into the differences in Jet reconstruction with use simulated of the FTK.

Methodology

To examine potential differences in track reconstruction from the use of FTK, requires the use of specially developed MC simulations. These MC samples are simulations of fully-leptonic and semi-leptonic $t\bar{t}$ pair-production events (see Section 1.3.1) generated by POWHEG and PYTHA (see Section 3.1) and contain the following special collections

- *HLT*: Jets reconstructed with tracks and vertices which are found by simulation of current ATLAS HLT tracking capabilities. Tracks are found only in RoIs with "fast tracking" for b-jet chains;
- *Offline*: Jets reconstruction with tracks and vertices which are found using the current algorithms for offline analysis.
- *FTK*: Jets reconstructed using tracks and vertices found by simulation of FTK in the ATLAS trigger. Track-finding “full-scan”: tracks are found globally, not just in RoI;
- *FTKO*: Jets reconstructed using tracks and vertices found by simulation of offline analysis using FTK. Track-finding “full-scan”: tracks are found globally, not just in RoI;

A singularly useful variable that can illuminate differences between these reconstruction collections is $R_{p_{\mathrm{T}}}$ of a jet [171], defined as the sum of transverse momentum for only those tracks that are both matched to a jet and associated to the PV as a portion of the sum of transverse momentum for all tracks matched to a jet:

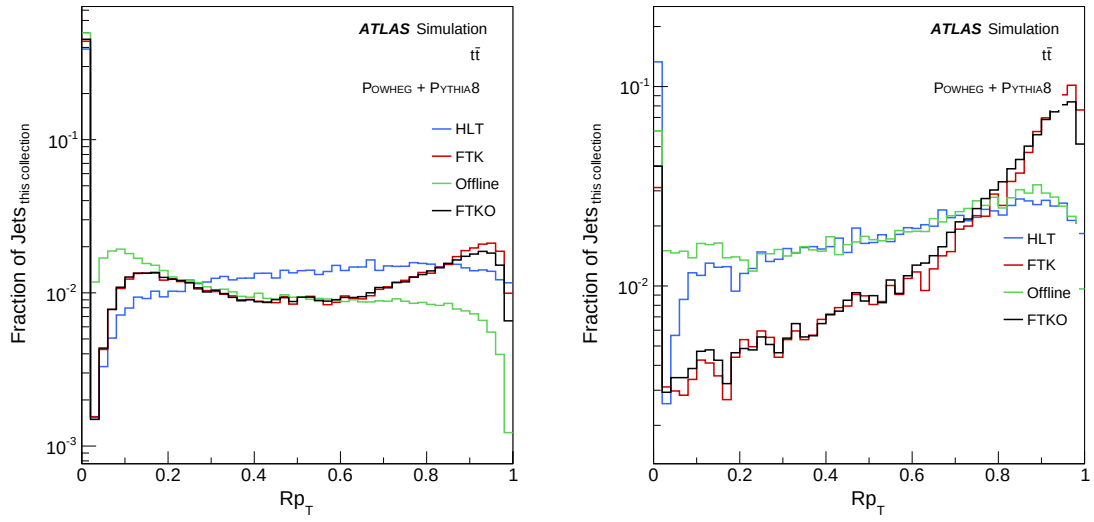
$$R_{p_{\mathrm{T}}} = \frac{\sum {p_{\mathrm{T}}}_{track}^{jetandpv}}{\sum {p_{\mathrm{T}}}_{track}^{jet}} \tag{5.1}$$

$R_{p_{\mathrm{T}}}$ is used as one of the key variables for developing the JVT BDT (see Section 4.3.3). In this analysis two objects are defined as matched when the distance between them in η-ϕ space is $\Delta R < 0.4$, as defined in Equation 2.10. $R_{p_{\mathrm{T}}}$ shows a physical estimation of how much of a jet's energy comes from the PV of the event. Selection requirements have also been placed on the tracks, such that they required $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ and $\eta < 2.5$, and a requirement on jets, such that they need $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and at least one track matched to them. Any other tracks or jets are not used in this analysis.

5.2.1 Results

$R_{p_{\mathrm{T}}}$

Looking at the results for $R_{p_{\mathrm{T}}}$, shown in Figure 5.2, there are sections of the distribution with strong agreement between jets from all of the collections, and some key areas of difference.



(a) All jets passing selections

(b) Lead jets only

Figure 5.2: R_{p_T} for tracks, jets and vertices from four different collections: FTK, FTKO, HLT and offline.

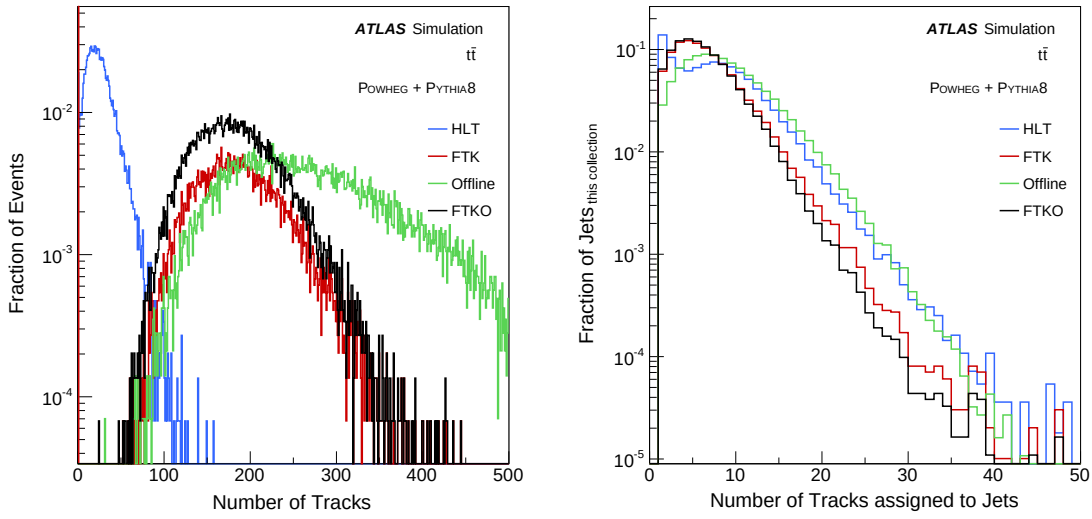
The FTK and FTKO collections show very close agreement, both when looking at all jets and when simply looking at the leading jet (with the highest p_T) of each event. All collections show a strong peak at zero indicating the majority of tracks matched to these jet are not also associated to the PV. When looking at results for all jets it can be observed that, in terms of shape, HLT shows the most difference to other collections. There is a gentle rise up to about 0.8, where it peaks, followed by a steeper fall. This indicates jets with a fairly high portion of their track-momentum coming from the PV are most common. The FTK, FTKO and offline collections have more in common, each showing an early peak with a drop off. This indicates these jet reconstruction methods found a high portion of track-momenta not from the PV. FTK and FTKO both have a second, more prominent peak close to unity, indicating that these jets are most likely to have close to all of their track-momenta sourced from the PV in our events.

When looking at results for leading jets, it can be seen that generally, jets reconstructed from all collections are more likely to have a high portion of their track-momenta originating from the PV of the event. This is most common for those from FTK and FTKO collections, which show very similar results to one another. This is less common for the HLT and offline collection.

To learn more about the origins of the differences in these results, and more about track-finding within jet reconstruction, exploration of some of the constituent elements of R_{p_T} were undertaken in more detail.

Number of Tracks

The first component of R_{p_T} explored was the number of tracks found within events. In Figure 5.3(a) the total number of tracks passing these selections is displayed. It can be seen that within the HLT collection far fewer tracks were found than in other collections. The number of



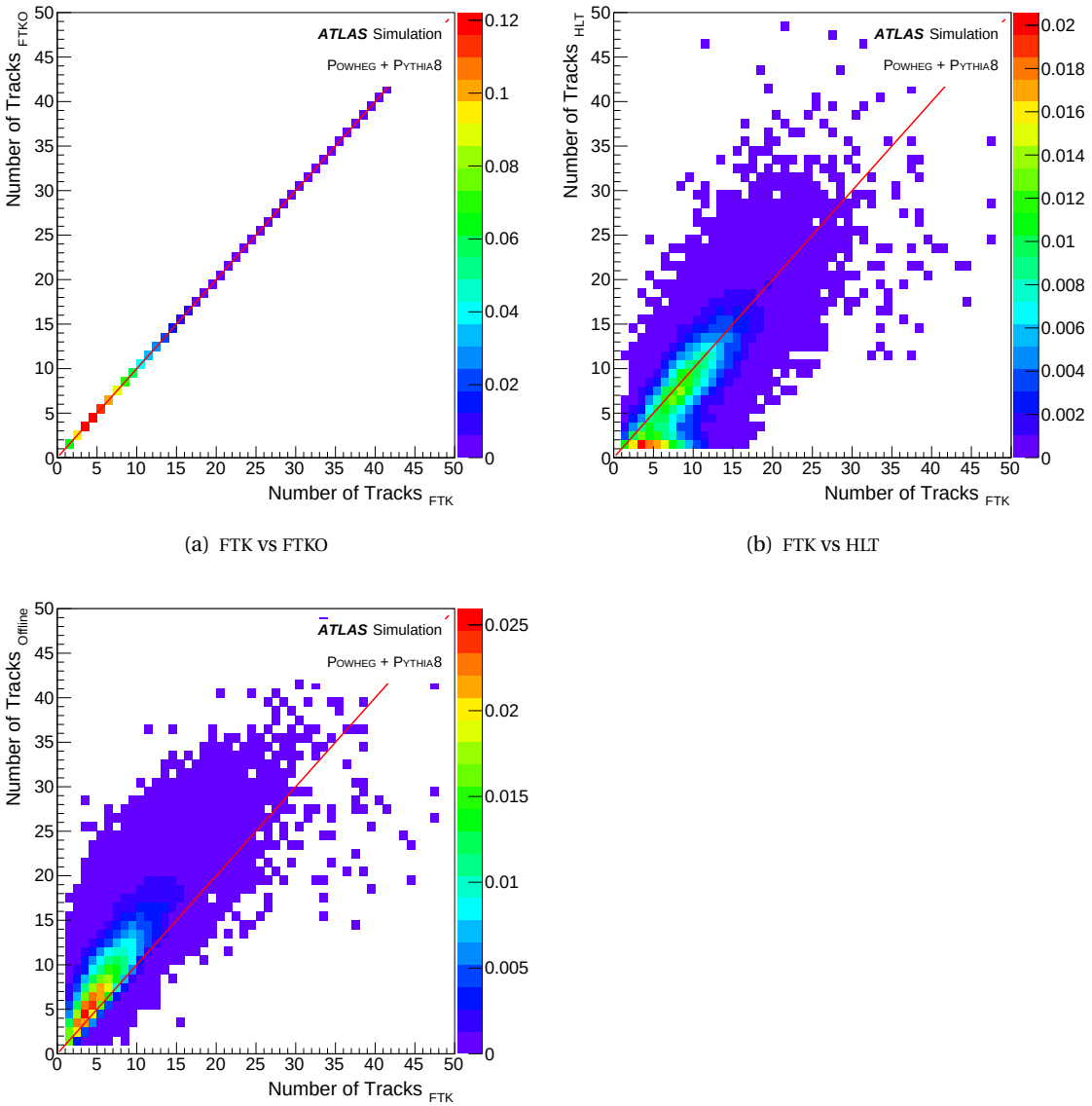
(a) (b)

Figure 5.3: The number of tracks found in each collection, (a), and the number of tracks matched to jets in each collection, (b).

tracks in HLT peaks at close to twenty tracks in each event. This is comparible to results taken from data, where the number of tracks (following a loose object-selection) per event (following an event-selection) is approximately 17.8 [196]. FTK and FTKO find many more, peaking close to 175 tracks per event – this difference is most likely due to the global nature of track finding. Offline jet reconstruction finds even more, peaking above 200 tracks per events, and continuing to a significant number of events finding at least 450 tracks per event. This is likely due to less time restrictions for offline track finding, allowing more tracks to be discovered.

Following this, the number of tracks matched to jets and how these vary for each collection was investigated. Looking at Figure 5.3(b) it can be seen that the majority of jets have up to seven tracks matched to them. HLT jets are most likely to have only a single track matched to them, the minimum requirement for this analysis. This peak quickly drops off into a dip, followed by a second, smaller peak close to seven tracks, before slowly dropping off. This means that, despite a high portion of tracks found with six tracks, or no tracks, there are a lower portion found between, with few tracks. For jets in FTK and FTKO something quite different is observed: again the majority of jets have fewer than eight tracks, but this peaks at about four, with many fewer jets having only one matched track. Following this peak the number of tracks matched to a jet drops off more quickly than for HLT jets. For offline jets it can be understood that there are a relatively small portion of jets with only one matched track, rising to peak at six. Following this the distribution drops off slowly, much like HLT.

Combining this information with what I have previously shown, these results suggest that the big difference in number of tracks between HLT and other collections in the region between 1 and 8 tracks comes from the regional, and not global, nature of track finding in HLT. Therefore, more jets are observed in FTK, FTKO and offline collections with a low to medium number of matched tracks, whereas for HLT these jets have closer to one matched track. This hypothesis



(a) FTK vs FTKO

(b) FTK vs HLT

(c) FTK vs offline

Figure 5.4: Jets can be matched one-to-one between different collections. Shown here is the number of tracks matched to FTK jets, where these jets have themselves been matched one-to-one with jets from other collections. The number of tracks matched to the respective matched jet is also shown.

was explored in more detail by comparing matched jets.

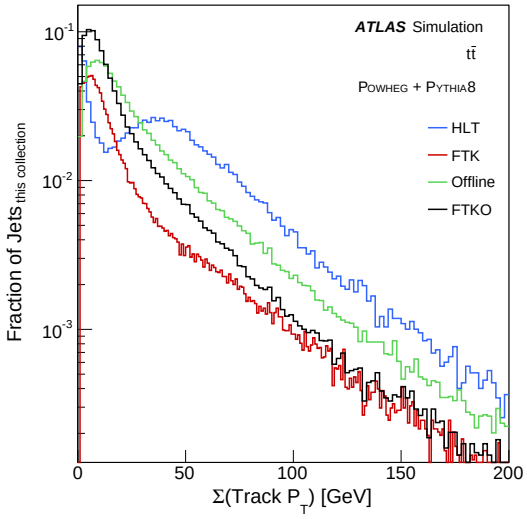
Ensuring a jet within the FTK collection is matched within a distance of $\Delta R < 0.4$ to a jet in the HLT, FTKO and offline collections allows comparison of the number of tracks matched to these jets one-for-one. Observe in Figure 5.4 how the number of matched tracks for these matched jets compare to jets from the FTK collection. Firstly, I present the results for FTK vs FTKO. Each of these pairs of matched jets have the same number of tracks matched within FTK as FTKO, an exact one-to-one result. It is visible that the majority of these jets have three or four matched tracks, with very few having more than ten. This reflects the previous result. Moving onto the comparison between HLT and FTK, it can be seen that there is a strong correlation between the two, but with two slight differences. At high numbers of tracks there slightly fewer matched tracks in FTK than HLT, though this is the minority of jets. In a much larger portion of these jets – those with very few tracks matched – there is a significant difference: the majority of the matched jet-pairs have a jet from the HLT collection with only a single track, and a jet from the FTK collection with three or four tracks. This skew towards a higher number of tracks found in FTK is present in all jets with relatively few tracks found in HLT. This matches the hypothesis from previous results and indicates that track finding for jets using the FTK is more likely to find several tracks where track finding in HLT has only found a low number.

When comparing jets from the offline collection to those in FTK, it can be once again seen that there is a reasonably good agreement between the numbers of tracks assigned to jets; however, for each matched jet pair there is likely to be a higher number of tracks in the offline case. This would indicate a general trend of better offline track finding. From explorations took place to see if there is a certain quality that these missed tracks have in common.

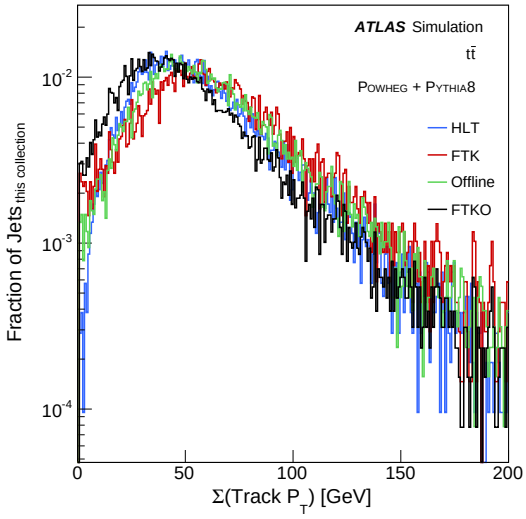
Sum of Track p_{T}

To a develop a broader understanding of the tracks in these jets I next constructed a larger portion of the original $R_{p_{\mathrm{T}}}$ variable, by looking at the denominator: the sum of p_{T} from all tracks matched to a jet, labelled $\Sigma p_{T\,Track}$. In Figure 5.5 I present the results for this variable.

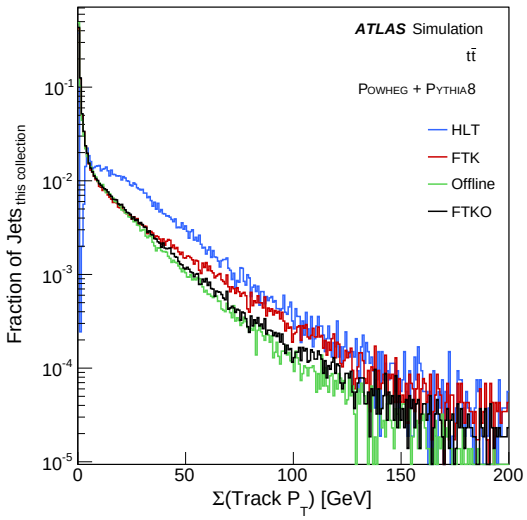
For the most part the results reflect those from the number of tracks matched to jets and the distributions follow a similar pattern. The sum of track p_{T} for jets in the FTKO, FTK and offline collections rise quickly to have a single peak - at 4 GeV, 6 GeV and 8 GeV respectively - and steadily drop off. The sum of track p_{T} for Jets in the HLT collection has its first, highest peak between 0 GeV and 2 GeV, before dipping, with a second, smaller peak at 34 GeV. The combination of HLT jets having, firstly, many fewer results with low, non-zero values than the other collections followed by a second peak at higher values of summed track p_{T} combined with secondlt, fewer tracks being found indicates that the tracks missed by jet-track finding in HLT are low p_{T} tracks. These results also indicate that the higher value of summed track p_{T} for the offline collection, compared to FTK, originates from the fact that more tracks are found in offline jets, increasing Sum of Track p_{T} for each jet. The differences in collection are not as apparent when looking at the lead jet (with the greatest value of p_{T}) in any event: distribution



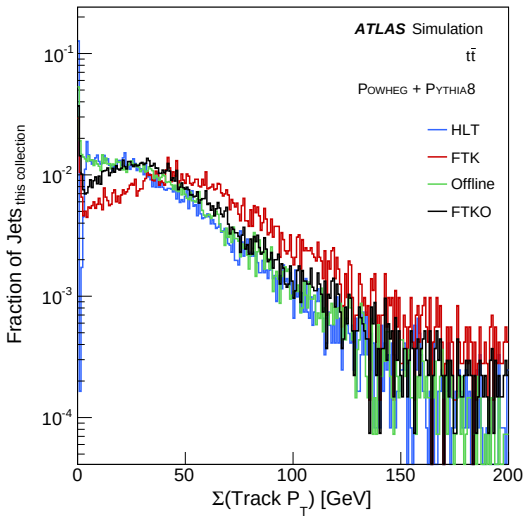
(a) Tracks matched to jets



(b) Tracks matched to lead jets



(c) Tracks associated to the PV and matched to jets



(d) Tracks associated to the PV and matched to lead jets

Figure 5.5: Once tracks are matched to jets, their p_{T} can be summed. This is shown for tracks matched all jets passing selection, (a), tracks matched lead jets, (b), tracks that are assciated to the PV and matched to jets, (c), and tracks that are associated to the PV and matched lead jets, (d).

shapes are much more similar; however, I see that in the HLT collection there is once again a scarcity of low-p_{T} results.

Following this, I also looked to the numerator of the $R_{p_{\mathrm{T}}}$ ratio, by exploring the summed momenta of tracks that are both matched to a jet and associated to the PV. Once more it can be observed that there is a very familiar difference between jets in the HLT collection, and strong similarities between jets from the FTK, FTKO and offline collections. A significant drop in the number of low-sum-p_{T} results for HLT jets is present in the results for these track sets, and those of lead jets. Overall, this indicates that track finding in FTK has similar capabilities for finding low-p_{T} tracks as in offline processing.

As with the number of tracks matched to jets from these collections, I undertakook a one-to-one comparison for matched jets. In Figure 5.6 it is shown that the sum of track p_{T} contrasts between pairs of jets from different collections.

It is shown that results from FTK and FTKO match exactly, with the majority of jets having a value between 2 GeV and 8 GeV. When compared to jets from the HLT collection, there is reasonably strong positive correlation; however, the majority of jets from the HLT collection have a value between 2 GeV and 6 GeV and are matched to jets from the FTK collection with a notably higher sum of track p_{T}. This would support the hypothesis that track-finding in the HLT is missing many low-p_{T} tracks that are found within the FTK collection. When inspecting the comparison between jet pairs from the FTK and offline collections, many cases can be seen where there is a higher value of offline jets when the value for FTK is particularly low (below around 20 GeV), otherwise there is a reasonably good correlation. This indicates the offline case is somewhat better at track-finding, following from offline jets having more tracks matched to them.

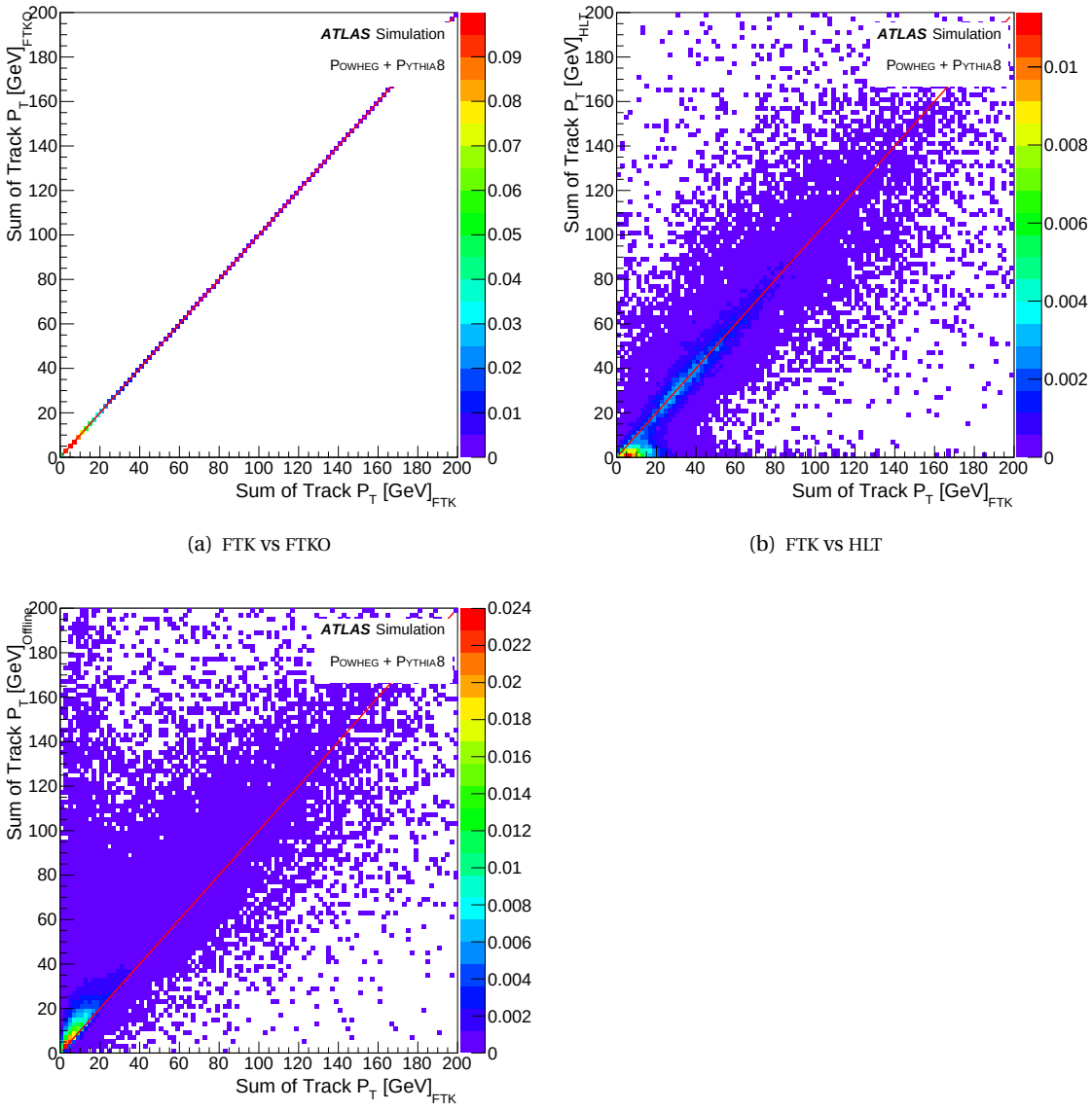
Track Width

From here I shall display the results from another insightful variable, track width, defined as

$$W_{Track} = \frac{\sum (\mathrm{Track} p_{\mathrm{T}}) \times (\Delta r_{fromjet})}{\sum \mathrm{Track} p_{\mathrm{T}}} \tag{5.2}$$

Looking at Figure 5.7 there are some clear difference between the different collections. First note, that the matching distance limits possible track width at a maximum of 0.4, as expected. For jets from the FTK, FTKO and offline collections there is a very similar distribution, with the portion of jets increasing steadily to a peak at a track width value of 0.24, before quickly dropping off as the limit of 0.4 is approached. For jets from the HLT collection, however, there is a very different distribution with a sharp rise up to a peak at track width of 0.08, followed by a sharp drop and a small second peak above 2.8. This seems to indicate a tighter track clustering about the jets in the HLT collection.

Investigating track width for matched jet pairs from different collections demonstrates that jets from FTK and FTKO are an exact match once again. When comparing FTK jets to those matched in HLT it can be seen that for FTK jets with values of track width between 0.02 and

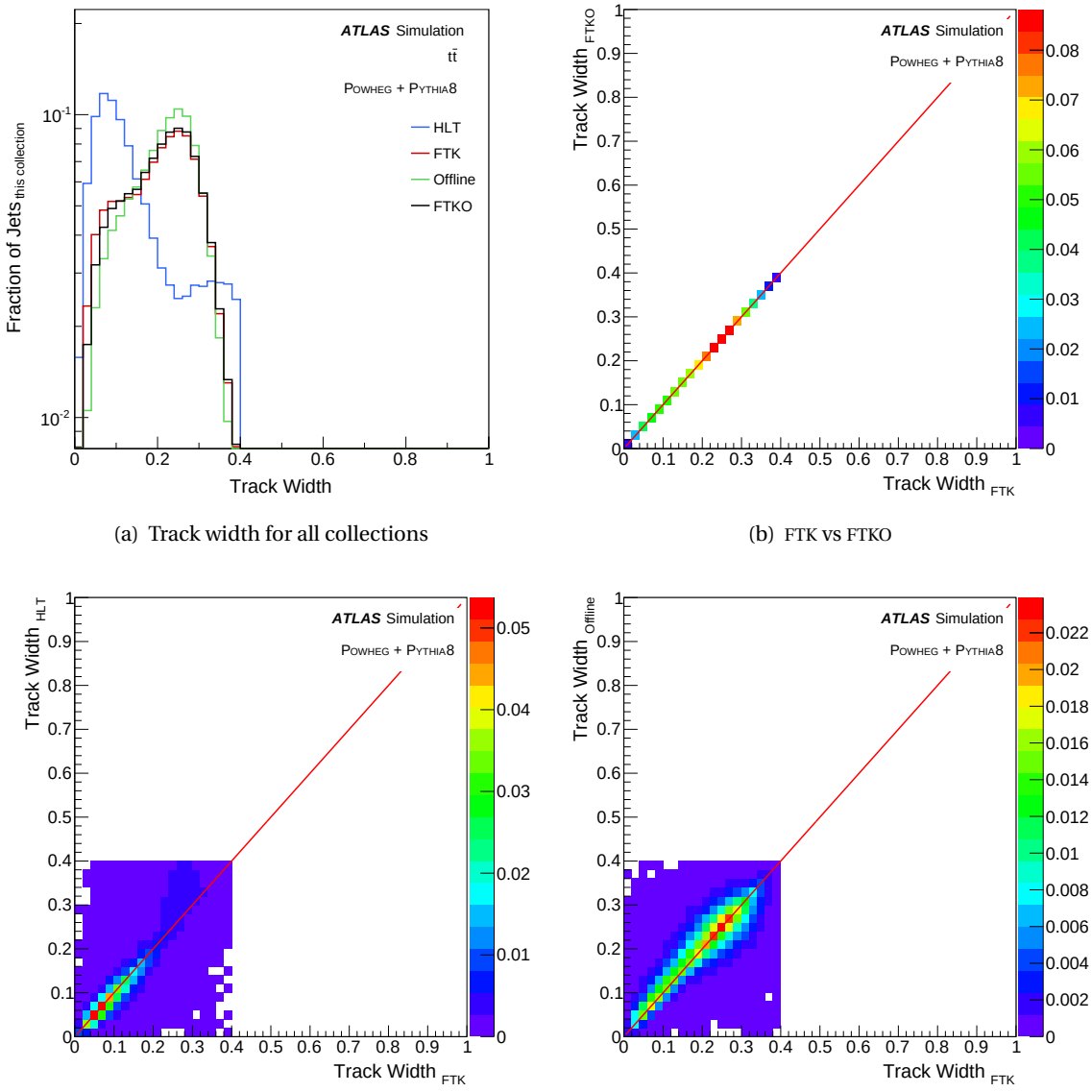


(a) FTK vs FTKO

(b) FTK vs HLT

(c) FTK vs offline

Figure 5.6: Jets can be matched one-to-one between different collections. Shown here is the sum of track p_{T}, Σ(Track p_{T}), for tracks matched to FTK jets, where these jets have themselves been matched one-to-one with jets from other collections. The sum of track p_{T} for tracks matched to the respective matched jet is also shown.



(a) Track width for all collections

(b) FTK vs FTKO

(c) FTK vs HLT

(d) FTK vs offline

Figure 5.7: Track width (Equation 5.2) is shown. Jets can be matched one-to-one between different collections. Also shown here is the track width for tracks matched to FTK jets, where these jets have themselves been matched one-to-one with jets from other collections. The track width of tracks matched to the respective matched jet is also shown.

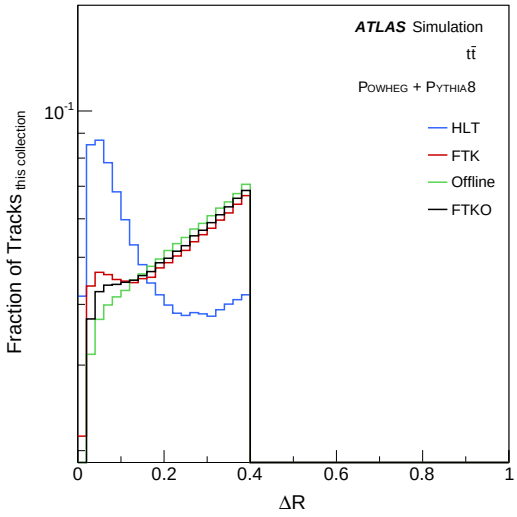
0.12 the value is skewed lower for the matched HLT jet, related to the peak in low values for HLT jets. At higher values track width increases more quickly for HLT jets than those from the FTK collection, with a clustering around 0.22 to 0.3, related to the FTK peak. Comparison of FTK jets and offline jets demonstrated that there is a very strong correlation in track width, as indicated, and a clustering close to where the peak is. Further investigations were then undertaken into the differences seen in track width. This was done with the use of related variables, shown in the following sections.

ΔR and Δz

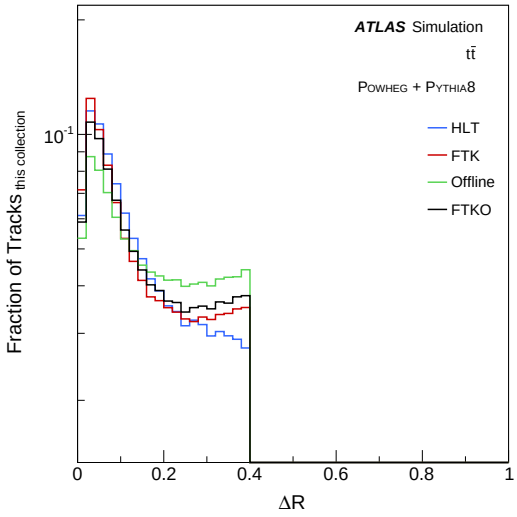
To investigate further potential differences in the distributions of tracks within jets the variable ΔR was investigated. This began with an analysis of the distances between these tracks and the centre of the jet they are matched to. Looking at Figure 5.8, the first thing to note is the hard cut-off caused by the maximum matching distance. Secondly, it is seen that the distributions in FTK, FTKO and offline jets are very similar to one another, and all very different to jets from the HLT collection. Jets from the FTK, FTKO and offline collection display a gradual rising in the number of tracks found when increasing the distance from the jet, up to a peak at the cut-off distance. This is likely related to the increasing likelihood of finding a track as the area searched increases quadratically. For jets in the HLT collection a very different pattern emerges: with a tight clustering of tracks about the jet centre, peaking at a distance of $\Delta R = 0.04$ before rapidly dropping off, with a second, much smaller peak at the cut-off distance. Looking at results for the lead jets in events it can be seen that this difference is not present, as jets in all collections find many more jets clustered about the jet centre. This increase in track-clustering is expected about high-p_{T} jets, as they tend to be more collimated. In all likelihood, the differences in track finding for HLT jets come from the regional nature of track-finding, with tracks from high-p_{T} jets prioritised. Global track-finding with the FTK means that many more tracks further from the jet centre are found, these are likely the same low-p_{T} tracks that are missing from HLT jets in the previous results.

Following this, this result was cross-checked with tracks that are both matched to jets and that are also associated with the PV of the event. For jets from the FTK, FTKO and offline collections, some evidence of clustering about the jet centre is shown, a result that is more prominent for leading jets. This is possibly related to the fact that tracks coming from the PV are matched to higher-p_{T} jets, originating from the hard-scattering in the event, with other jets in the event sourced from soft-interactions and pile-up. For jets from the HLT collection a far more extreme clustering about the jet centre is observed, once again displaying these differences between the regional and global track-finding.

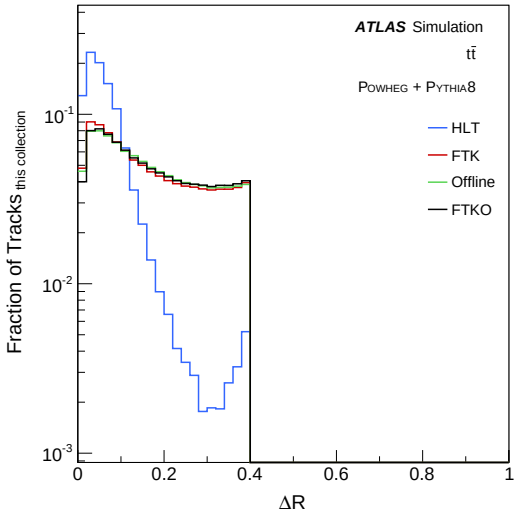
Further to this, the distribution of jet-tracks about the leading-track was inspected. Here the leading track is defined as the track with the highest-p_{T} that is also matched to a jet. In Figure 5.9 these results are shown. It is clear that the distribution for tracks from the FTK, FTKO and offline collections are very close, and that tracks from the HLT collection are slightly different. It is noted that all distributions follow a rise-and-fall curve, with tracks from the HLT collection



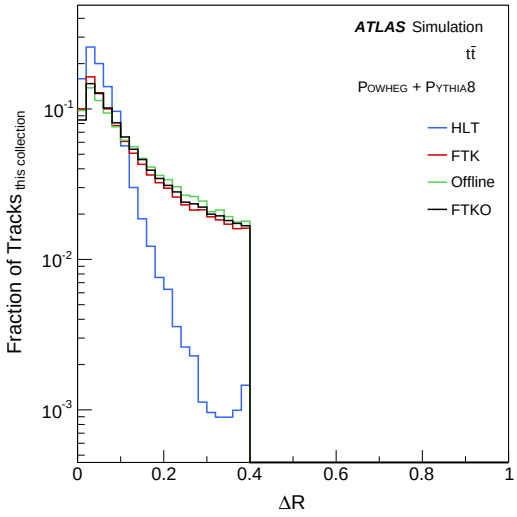
(a) Tracks matched to jets



(b) Tracks matched to lead jets

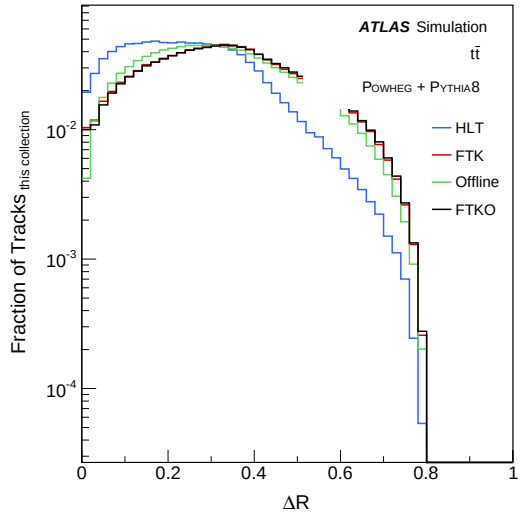


(c) Tracks associated to the PV and matched to jets

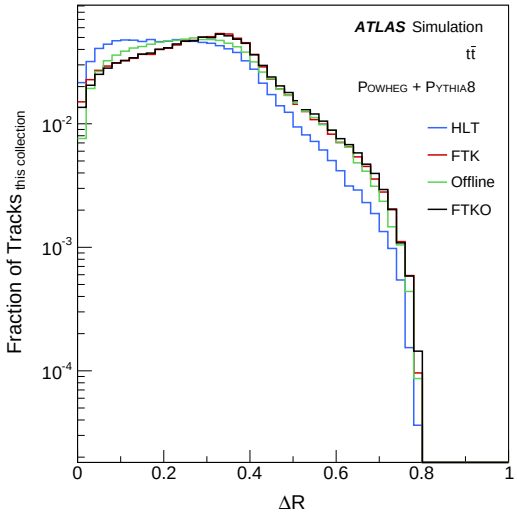


(d) Tracks associated to the PV and matched to lead jets

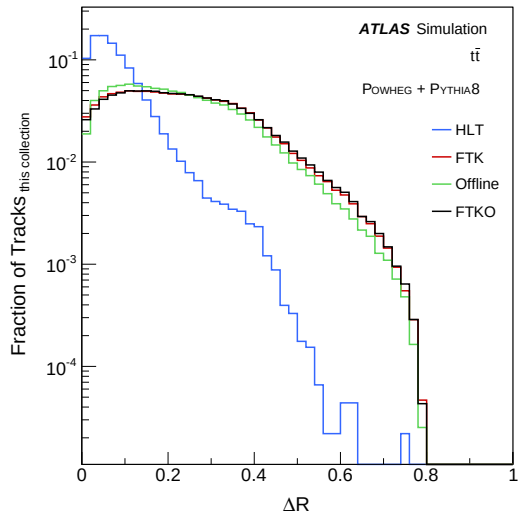
Figure 5.8: ΔR between tracks and the jet they are matched to. This is shown for tracks matched all jets passing selection, (a), tracks matched lead jets, (b), tracks that are assciated to the PV and matched to jets, (c), and tracks that are associated to the PV and matched lead jets, (d).



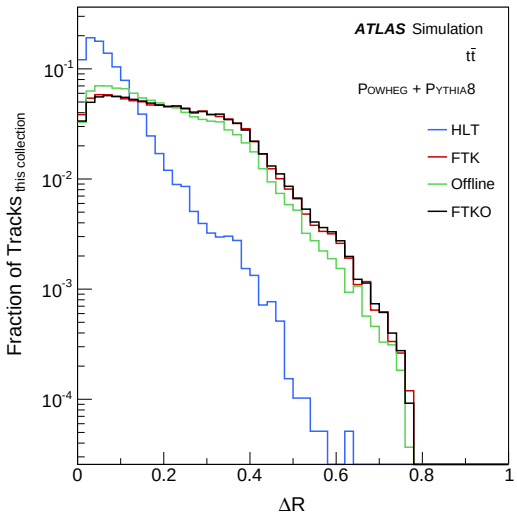
(a) Tracks matched to jets



(b) Tracks matched to lead jets



(c) Tracks associated to the PV and matched to jets



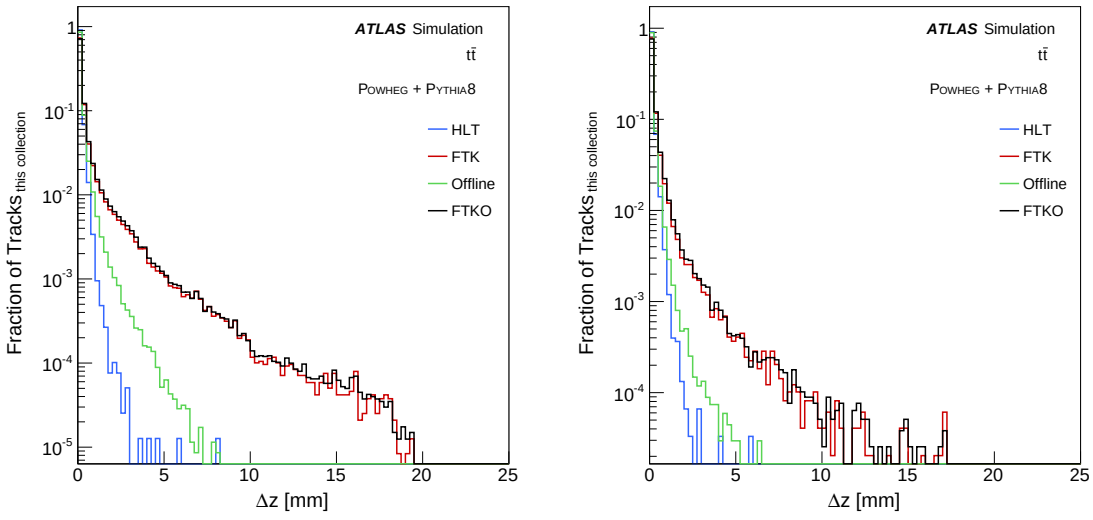
(d) Tracks associated to the PV and matched to lead jets

Figure 5.9: ΔR between tracks and leading track matched the same jet. This is shown for tracks matched all jets passing selection, (a), tracks matched lead jets, (b), tracks that are assciated to the PV and matched to jets, (c), and tracks that are associated to the PV and matched lead jets, (d).

peaking earlier than other distributions. The gradual increase up to the first peak is related to the quadratically increasing area about the lead track in which other tracks can be found. The effects of the selection, $\Delta R < 0.4$ has two effects on results: the first is a hard cut-off at the maximum possible distance between two tracks matched to the same jet, $\Delta R = 0.8$, the second is that as $\Delta R = 0.4$ is reached, the maximum possible distance for tracks very close to the jet centre is approached. The further the distance moves from 0.4 towards 0.8 the fewer possible locations there are to locate pairs of tracks this distance apart that are both still matched to the jet. This likely contributes significantly to the increasingly fast drop off following the crest of the curve. When looking at tracks matched to leading jets or tracks matched to all jets it can be seen that those from the HLT collection have an earlier peak than other collections. Also demonstrated is that track distances from the FTK, FTKO and offline collections have a slightly more prominent peak

Once again, this was cross-checked by only including tracks that are both matched to a jet and associated to the PV of the event. For tracks from the FTK, FTKO and offline collections there is a very similar distribution to those tracks only matched to jets; however, for tracks in the HLT collection there is now a very prominent group of tracks close to the lead track. This clustering about the lead track is likely due to the collimated nature of high-p_{T} jets originating from the PV, caused by the hard-scattering event. These high-p_{T} jets have central lead tracks and dominate the results for HLT in a way that global track finding in FTK is less susceptible to.

The final variable examined to investigate these differences was the longitudinal distribution of tracks that are matched to jets, about the PV, Δz. The results for this variable are shown in Figure 5.10.



(a) Tracks matched to jets.

(b) Tracks matched to lead jets.

Figure 5.10: Δz between tracks and the PV. This is shown for tracks matched to all jets, (a), and tracks matched to lead jets, (b).

It can observed that tracks found from the HLT collection are clustered the most closely about the PV, with 0.1 % of tracks reaching 2.5 mm away. Tracks from the FTK and FTKO col-

lection are the least tightly clustered, with 0.1 % reaching as far as 10 mm. Offline tracks lay in between, with 0.1 % reaching 5 mm distance from the PV. This pattern is the same when looking at tracks only matched to lead jets. The greater spread of tracks from the FTK and FTKO collection once again demonstrates differences between regional and global track-finding in the ATLAS experiment.

5.2.2 Conclusion

Through this analysis I undertook an exploration of the differences in track-finding between use of the ATLAS HLT, FTK, FTKO and offline reconstruction, to build solid ground work through investigation of several informative variables. These results lead to a clear, if preliminary, understanding of some key similarities and differences between the use of tracks with the FTK compared to other methods. It was shown that the tighter clustering about the matched-jet centre, lead track and PV in the tracks from the HLT collection is not present in tracks from the FTK collection, and propose this is due to the global nature of track-finding. This result is much closer to results seen from the more detailed process of offline track-finding in events, and leads to benefits, such as more low-p_{T} tracks being found in FTK events than HLT events. These differences are likely to improve tracking with jets, ultimately improving the efficiency and reliability of jet reconstruction in ATLAS.

6 SMALL JET RECONSTRUCTION WITH SOFT DROP GROOMING

To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science.

Albert Einstein

6.1 Analysis Overview

Since the discovery of the Higgs boson, a key area in the search for new physics is investigating the Yukawa coupling, in particular that of the heaviest SM particle, the top quark. The $t\bar{t}H$ channel of Higgs production involves a direct coupling between the top and the Higgs, and therefore presents a unique opportunity for study. The $H \to b\bar{b}$ decay channel has the highest Branching Ratio (BR), at 58 %, of all decays for a SM Higgs of 125 GeV (see Section 1.3.1). Searches for signals for this particular decay are complicated by the most common background process within the LHC, multi-jet production and top quark pair production, $t\bar{t}$, with additional hadronic jets. The top-quark decay with the highest BR is $t \to bW$, at 99.8 %. Jets from b-quarks, therefore, play a key role in the search for understanding the Higgs boson. The subsequent decay of the produced W bosons can lead to either a 2-jet, 1-jet or 0-jet state with a BR of 45.7 %, 43.8 % and 10.5 % respectively. Thus, accurately identifying the sources of jets is paramount, but complex. Work within hadronic jet identification and classification, however, can improve understanding of the different physics taking place in these events. This may ultimately lead to better separation between events containing signals of physics of interest, such as events containing the Higgs boson, from background events.

Increases in the luminosity of the ATLAS experiment, both ongoing and planned for the future, will lead to some detriment in jet reconstruction, most particularly due to pile-up. Understanding the internal structure of jets in events at ATLAS through the utilisation of jet substructure (JSS) observables can significantly improve the understanding of physical processes

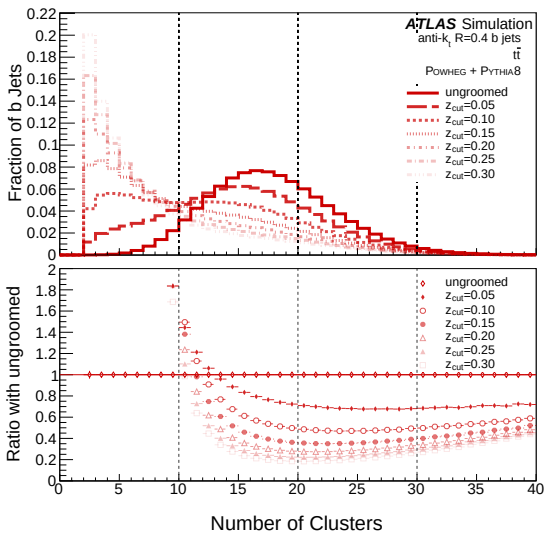


Figure 6.1: NCLUS for the b-flavour-tagged jets from the $t\bar{t}$ MC sample at each stage of soft drop grooming, with a ratio to the ungroomed value. Applying soft drop grooming to the jets steadily reduces the value of NCLUS. This is dicussed in detail in Section 6.6.1.

taking place within events (see Section 4.5). It is imperative, therefore, that accurate reading of these JSS variables is undertaken. Doing so for key variables such as jet mass, critical in identifing boosted hadronically decaying particles, is complicated by these increasing levels of pile-up within the detector. Accurately identifying jet mass is complicated by the presence of Non-Global Logarithms (NGL) (Section 4.7), an issue that is exacerbated by these higher levels of pile-up.

Soft drop grooming is a jet grooming procedure, used to remove soft and wide-angle radiation from the jet, that is formally insensitive to NGL (see Section 4.7). The calculation of the masses of jets that have the soft drop procedure applied is, therefore, insensitive to these NGL. The motivation for this analysis is to ascertain whether applying soft drop grooming to flavour-tagged small jets reconstructed within ATLAS can be beneficial for discrimination between $t\bar{t}H$ signal events and events of the most prominent backgrounds, such as $t\bar{t}$. To do this I undertake an investigation of JSS observables from reconstructed flavour-tagged small jets of different origin with different strengths of soft drop grooming applied, sourced from both MC simulations and ATLAS data.

In this chapter, I construct the focus of the primary analysis: investigating the effects and possible benefits of applying soft drop grooming to flavour-tagged small jets reconstructed within ATLAS. Following this, I explore the results of the studies into soft drop grooming on flavour-tagged small jets from both MC simulation samples and ATLAS data. Firstly, in Section 6.6 I present the investigation of effects soft drop grooming has on reco-jets, particularly on the JSS observables NCLUS, and jet mass. The results of these investigations prompted a more detailed exploration of soft drop grooming with tau-jets, which is covered in Section 6.7. Here I show results relating to "lost" tau-jets and "real" vs "fake" tau-jets. The next section, Section 6.8, covers differences in quark- and gluon-jets when subjected to soft drop grooming.

Following this, I look at the effects on b -jets that have been also identified through traditional b-tagging methods in Section 6.9. After this I introduce the results from flavour-tagged small jets originating in $t\bar{t}H$ MC samples in Section 6.10. These will be compared to the results from $t\bar{t}$ MC samples. Then I move onto studying the results of applying soft drop grooming to truth-particle jets, and how these compare to reco-jets in Section 6.11. Section 6.12 will investigate the effects of soft drop grooming on jet response. Finally, I will cover the analysis and results from utilising data from the ATLAS detector in Section 6.13.

6.2 Data and Monte Carlo Samples

Each of the samples used in these studies, both collected by the ATLAS detector or MC simulation produced as part of the official ATLAS MC production, have been chosen such that samples for comparison closely match one another in terms of their production, format and for specific event- and object- selection required (such as the necessary calorimeter topocluster information).

An important factor in choosing the best MC samples for comparisons within the analysis is the distribution and average value of pile-up found within the events, shown in Figure 6.4. Where comparison of results are made between MC samples and ATLAS data, or either between different MC simulation samples, matching weighted pile-up distributions are used, as shown in Figure 6.4(b).

6.2.1 ATLAS Data

Some of the analyses presented here utilise data from proton-proton collisions, at $\sqrt{s}$ =13 TeV centre-of-mass energy, collected by the ATLAS detector during 2018. This data has to pass the data quality criteria known as the Good Runs List (GRL) – requiring that beam of the LHC is stable, that all of the subdetectors within ATLAS have correct operating voltages, and that the magnetic fields of the toroidal and solenoidal magnets are of the correct strength [197]. Further validation is also undertaken by the ATLAS Data Quality group, studying data for deviations with respect to previously gathered, and well-understood runs. 41.9 fb^{-1} of 58.5 fb^{-1} "Good for Physics" data from 2018 (See Figure 6.2). This subset of the ATLAS data was chosen for its specific format and for the required pre-selection on events and objects necessary to undertake the analysis, such as inclusion of necessary calorimeter cluster information.

It should be noted that the pile-up distribution of the final sub-set of ATLAS data analysed in this thesis does not match exactly with the full 2018 data. This was deemed acceptable for these performance studies, where the physics of the reconstructed hadronic jets, and their behavious under soft drop grooming, is the focus.

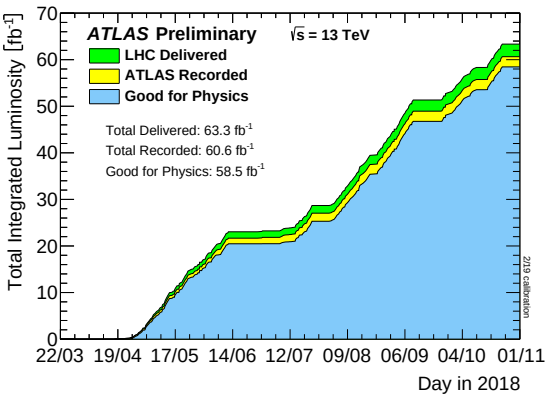


Figure 6.2: Cumulative luminosity versus time delivered to ATLAS by the LHC (green), recorded by ATLAS (yellow), and certified to be quality data that is "Good for Physics" (blue) during stable beams for proton-proton collisions at 13 TeV centre-of-mass energy in 2018 [7].

6.2.2 Monte Carlo Simulation Samples

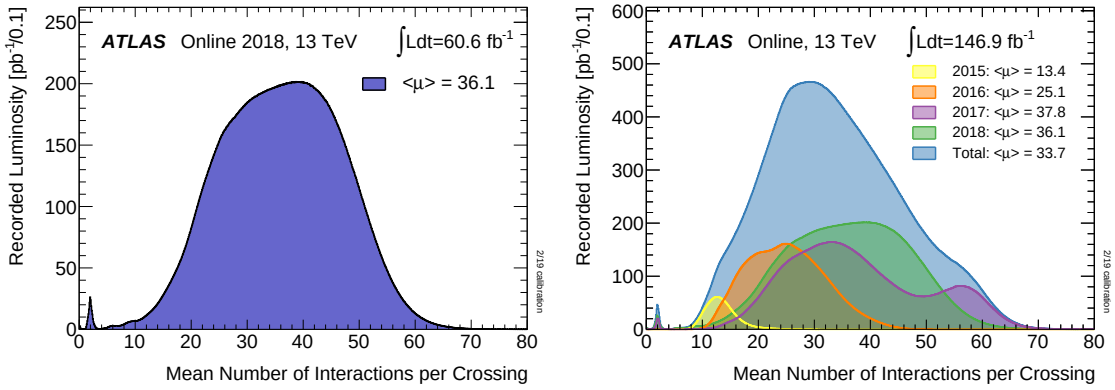
These analyses rely on samples of MC simulations produced as part of the official ATLAS MC production, discussed in Section 3.1. The first of these used is a sample of (leptonic and semi-leptonic) $t\bar{t}$ MC events simulated using POWHEG-BOX 2 [198] and utilising PYTHIA [119] for modelling of PS, hadronisation and UE. The second of the MC simulation samples utilised is semi-leptonic $t\bar{t}H$ events (see Section 1.3.1). This sample is modelled using the same generators: as above. Both of these samples have weighted pile-up distributions to match the pile-up conditions from ATLAS in 2017 (see Figures 6.3(b) and 6.4(a)).

Through the latter section of analysis, in which I make comparisons to ATLAS data from 2018, I utilise two different MC simulation samples. One contains events with $t\bar{t}$ samples simulated using POWHEG-BOX 2 and PYTHIA, which is referred to as "$t\bar{t}_2$". Also shown is results results from $t\bar{t}_2$ following application of the selection process described in Section 6.3.2, which is labelled "$t\bar{t}_2 Sel$". The second contains dijet samples, simulated by SHERPA [199]. Each of these have weighted pile-up distributions to match conditions from ATLAS in 2018, shown in Figure 6.4(b).

6.3 Selection Processes

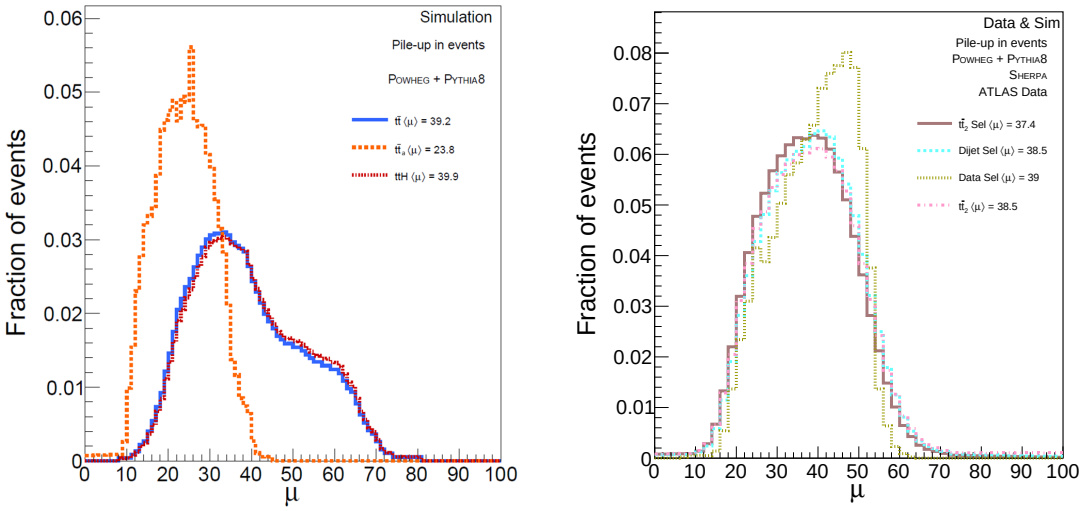
6.3.1 Object Selection

The first physics objects I make use of in this analysis are *calorimeter topoclusters* (from now on referred to as *clusters* or *topoclusters*), which I use to reconstruct *calo-jets* (from now on referred to as *reco-jets*). I select only those clusters with positive energy values and implement the FastJet package to reconstruct jets with the Anti-k_T clustering algorithm, previously covered in Section 4.2. I reconstruct the jets with a clustering radius of $R = 0.4$, and require that all jets used for analysis possess at least two clusters.



(a) 2018 (b) 2015 - 2018

Figure 6.3: Luminosity-weighted distribution of the mean number of interactions per crossing for proton-proton collisions during Run 2 at CoM energy, $\sqrt{s} = 13\,\text{TeV}$ data [8] for 2018 (a) and for 2015 - 2018 (b).



(a) 2015 and 2017 weighting. (b) 2018 weighting and data.

Figure 6.4: Pile-up distributions for the different MC samples, and data sample. Shown is number of particle interactions per bunch, μ.(a) shows MC simulation samples weighted to 2015 and 2017 data. (b) shows both MC samples weighted to 2018 data and the sub-set of ATLAS data used in comparison studies.

I also make use of final-state truth-particles in order to reconstruct *truth-particle jets*. These truth-particle jets are reconstructed from at least two final-state truth-particle "constituents", though this will not include any neutrino or muon particles, as these are rejected. Once again, I employ the FastJet package to reconstruct Anti-k_T jets with radius of $R = 0.4$.

The final three jet objects are all "out of box", and are not produced during my individual analysis, but taken directly from the MC and data samples. These are constructed using the reconstruction processes described in Section 3.3. The first of these jet types are *EM-jets*, which are reconstructed from uncalibrated topoclusters (therefore at "EM" scale, see Section 4.3.1).

These have been reconstructed using the Anti-k_T algorithm, with a radius $R = 0.4$. The next of these jet types are *truth-jets*, previously reconstructed from truth-particles using the Anti-k_T algorithm, with a radius $R = 0.4$. The final "out of the box" jet types I use in this analysis are *traditional tau-jets*. These are jets defined as originating from the hadronic decays of the τ lepton by traditional τ -tagging methods. In total, this gives five types of jet used within the analysis, reviewed in Table 6.1.

Jet Type	**Description**
Reco-jet	Reconstructed during my analysis from calorimeter topoclusters using FastJet and Anti-k_T at $R = 0.4$
Truth-particle jet	Reconstructed during my analysis from final state "truth-particles" using FastJet and Anti-k_T at $R = 0.4$
EM-jet	Previously reconstructed from uncalibrated calorimeter topoclusters using Anti-k_T at $R = 0.4$
Truth-jet	Previously reconstructed from final state "truth-particles" using Anti-k_T at $R = 0.4$
Traditional tau-jet	Jet identified as originating from τ by traditional tagging methods.

Table 6.1: The types of jet used in this analysis.

I place selections on all jets used within the analysis. Each of these jets must have $|\eta_{max}| = 2.5$ and $p_{T\,min} = 25\,\text{GeV}$. I also require that each reco-jet used is matched sufficiently close within the detector to at least one EM-jet, within a distance of $dR_{max} = 0.4$.

In studies only using MC simulated events I also require a matching of this fashion between the reco-jet, at least one reconstructed truth-particle jet and at least one truth-jet. I also require a p_T matching between the truth-particle jet and truth-jet, such that the two objects have $\Delta p_{T\,max} = 25\,\text{GeV}$. Any jets not meeting these requirements are discarded.

Some of these analyses make use of "out of the box" reconstructed electrons and muons (see Section 6.3.2).

6.3.2 Event Selection

Following the selection of the physics objects, I employ criteria to select only those events I wish to use. There are two event selection configurations I use in these analyses, the first of which will only be used for MC simulation studies, as it requires truth-information not present in ATLAS data. Here I ensure an event always contain at least one reco-jet with a matched EM-jet (both meeting the jet selection and matching requirements). They must also contain at least one truth-jet and at least one truth-particle jet, both matched to a reco-jet and to one another. This matching criteria is explained above.

Selection for Analyses with ATLAS Data

Data from the ATLAS detector does not contain any "truth" information about events. Therefore, studies that involve this ATLAS data require a common event selection to give an accurate

comparison. These selections are combined with further criteria used in traditional searches for $t\bar{t}$ events. This set of event selection criteria is used to select $t\bar{t}$ pair production events with leptonic and semi-leptonic decays [200].

Here, the only jets that are available are the "out of the box" EM-jets and the reco-jets I have constructed from at least two clusters, as defined in Section 6.3.1. As before, each of these jets must be central to the detector, $|\eta_{max}| = 2.5$, and have a minimum transverse momentum, $p_{\mathrm{T}\,min} = 25\,\mathrm{GeV}$. I also require that each reco-jet used is matched sufficiently close within the detector to at least one EM-jet, within a distance of $dR_{max} = 0.4$. Every event requires at least 4 of these jets, with either 1 or 2 of them b-tagged (see Section 6.4.3) Further to this each event must have exactly one electron or muon with a maximum transverse momentum, $p_{\mathrm{T}\,max} = 28\,\mathrm{GeV}$.

It is also required that minimum values are set for $E_{\mathrm{T}}^{\mathrm{miss}}$ and for the transverse mass of the W boson, m_T^W. For events with an electron these minimum values are set to $E_{\mathrm{T}}^{\mathrm{miss}} > 30\,\mathrm{GeV}$ and $m_T^W > 30\,\mathrm{GeV}$. For those events in which an muon is found a minimum for their combined value is required, $E_{\mathrm{T}}^{\mathrm{miss}} + m_T^W > 60\,\mathrm{GeV}$.

$$m_T^W = \sqrt{2 p_T^{\ell} E_{\mathrm{T}}^{\mathrm{miss}} 1 - \cos\phi} \tag{6.1}$$

6.3.3 Selection Cuts

For some studies in these analyses, I will use the application of one-dimensional selection cuts to further explore the physics taking place. Application of selection cuts on a single observable, such as rejecting all jets with a jet mass below a chosen maximum value, $m_{max} = 50\,\mathrm{GeV}$ for example, may have a stronger effect on jets of a particular flavour-tag, or from a particular sample. If used correctly, this can increase the "purity" of a sample, in regards to a chosen flavour. I will then explore whether using soft drop grooming can improve this selection process, allowing an increase in the maximum possible purity of the sample.

Following this, the application of these selection cuts can also be limited, such that a minimum chosen Working Point (WP) must be reached following a cut for it to be valid. The chosen WP used will be that 80 % of the chosen "signal" jets must remain followin a selection cut. These signal jets may be jets from a chosen flavour or from a given sample. Therefore, I look only at those cuts that will return at least 80 % of the chosen jet origin or flavour. Once again, I will which of these selection cuts return the best purity for my flavour or sample of choice. Through this method, I hope to investigate whether soft drop grooming can improve the maximum possible purity achieved by the selection cut.

6.4 Jet Flavour-Tagging

Following the reconstruction and selection of jets, the next stage is to assign flavour information, labelling the proposed origin of the jet. In these analyses, reco-jets are assigned flavour-

tags through three different methods, which can be applied individually or in combination. Each of these methods and their application in analysis are described below.

6.4.1 Truth-Flavour-Tagging

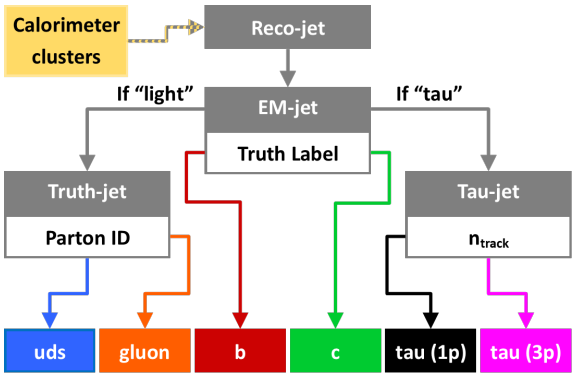


Figure 6.5: Schematic of truth-flavour tagging sequence used in this analysis. Also shown is the sub-categorisation of tau-flavour-tagged jets into 1- and 3-pronged tau-jets.

MC simulation samples contain truth information that can be utilised to determine the origin of jets through the process of *truth-flavour-tagging*. Truth-flavour-tagging is a process that involves matching a jet to a truth-particle within $dr_{max} = 0.3$. EM-jets from the chosen samples already have this information for this analysis to access. Any reco-jets matched to an EM-jet with a truth-flavour label of b, c, τ or "light" are kept, all others are either flavour-tagged as "lost taus" (See Section 6.4.2) or rejected.

In this context, light-jets are those which have been matched with light partons. Further truth information, taken from the truth-jets matched to these light-jets, can be used to determine whether the origin is a light-quark (u, d or s) or a gluon. This *PartonTruthLabelID* corresponds to the ID of the highest energy parton ghost-matched to the jet [166]. The jet can then be labelled as either a gluon- or uds-jet. A schematic detailing this truth-flavour-tagging process is shown in Figure 6.5.

6.4.2 Further Tau-Flavoured Jet-Tagging

1-Pronged and 3-Pronged Tau-Jets

Those reco-jets that have now been assigned a flavour-tag of "tau" are further sub-divided into *1-prong* and *3-prong* tau-jets. For this to be assigned, the reco-jet must be matched with a tau-jet found through traditional τ -tagging techniques. Any reco-jets flavour-tagged as tau-jets without this match are passed on to be "lost taus" (See Section 6.4.2). Next, I access the number of ID tracks associated to the matched tau-jet. This relates to the number of charged decay-products discussed in Section 3.3.4. A single track, or three tracks (the only two available number of tracks in this case), leads to the tau-flavoured reco-jet being categorised as a 1-

pronged tau reco-jet, or 3-pronged tau reco-jet respectively [146]. A schematic detailing this flavour-tagging process is shown in Figure 6.5.

Fake and Real Tau-Jets

Beyond classification of tau-flavoured jets by the number of prongs they possess, I further divide them into “fake” and “real” taus. I determine this by checking the tau-jet (found by traditional τ -tagging techniques) that is matched to a reco-jet now flavour-tagged as either a 1-pronged or 3-pronged tau. If this matched tau-jet is, itself, matched to a truth-particle defined as a stable, final-state, non-leptonically decaying τ -lepton, then my reco-jet is labelled as a “Real” 1- or 3-pronged tau, representing a hadronic τ decay within the MC simulation. Otherwise, the 1- or 3-pronged tau is labelled as “Fake”. A schematic detailing this flavour-tagging process is shown in Figure 6.6.

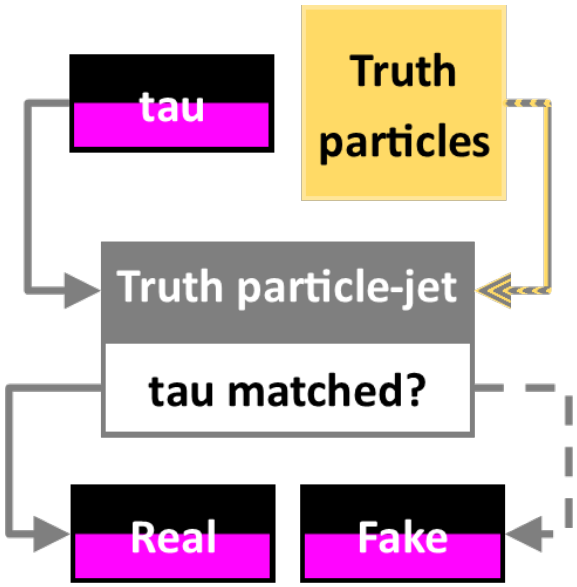


Figure 6.6: Schematic of process for flavour-tagging tau-flavoured reco-jets as either "Real" or "Fake".

"Lost" Tau-Jets

The final, additional, tau-related flavour-tag used in this analysis is that of “lost tau”. These are to account for those hadronic tau decays not identified by traditional τ -tagging methods.

There are two ways a reco-jet will be tagged as a lost tau-jet. Firstly, if the truth-flavour-label of the matched EM-jet is τ and the reco-jet is not matched to a tau-jet found through traditional τ -tagging methods, but is matched to a truth-particle jet that is a non-leptonically decaying τ-lepton. Secondly, if the truth-flavour-label of the matched EM-jet does not match that of light-, *b*, *c* or tau, but the reco-jet is matched to a truth-particle jet that is a non-leptonically decaying τ-lepton. A schematic detailing this flavour-tagging process is shown in Figure 6.7.

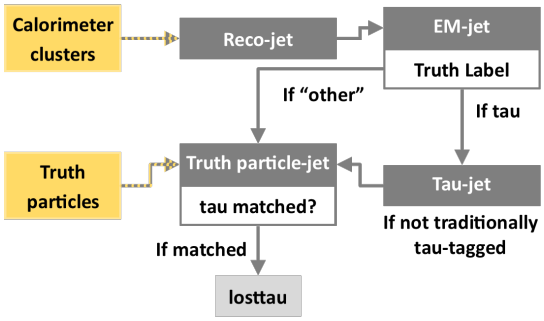


Figure 6.7: Schematic of process for flavour-tagging "lost" tau-flavoured reco-jets.

6.4.3 Applying b-Tagging

Additional to, or instead of, utilising truth-information for flavour-tagging the reco-jets, I also make use of traditional b-tagging methods. Each EM-jet has a b-tagging value assigned according to the MV2c10 algorithm described in Section 4.4. If this b-tag value passes the threshold of 0.9349 (see Table 4.1), then any reco-jet matched to said EM-jet will be labelled as "tagged". As no truth-level information is available in data from the ATLAS detector, this will be the primary flavour-tagging method used for those studies. It will also be used to further investigate those reco-jets already truth-flavour-tagged as b -jets. A schematic detailing this flavour-tagging process is shown in Figure 6.8. The efficiencies and rejection rate of this cut are shown in Section 4.4 and Table 4.1.

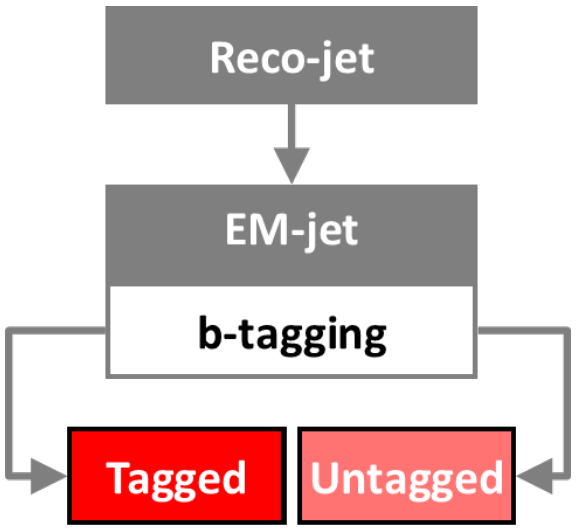


Figure 6.8: Schematic of process for applying traditional b-tagging to reco-jets using the the MV2c10 algorithm.

6.5 Application of Soft Drop Grooming

Following any strength of soft drop grooming, I require reco-jets to possess a minimum of two clusters, otherwise they are discarded. As truth-level information, and therefore truth-particle jets, do not exist in ATLAS data, those analyses using the data-tailored selection process de-

scribed in Section 6.3.2 will have this requirement only. Those analyses making use of truth-flavour information and the standard event selection process have the additional requirement that truth-particle jets must possess a minimum of two constituent truth-particles following the application of soft drop grooming, otherwise the reco-jet it is matched to is discarded.

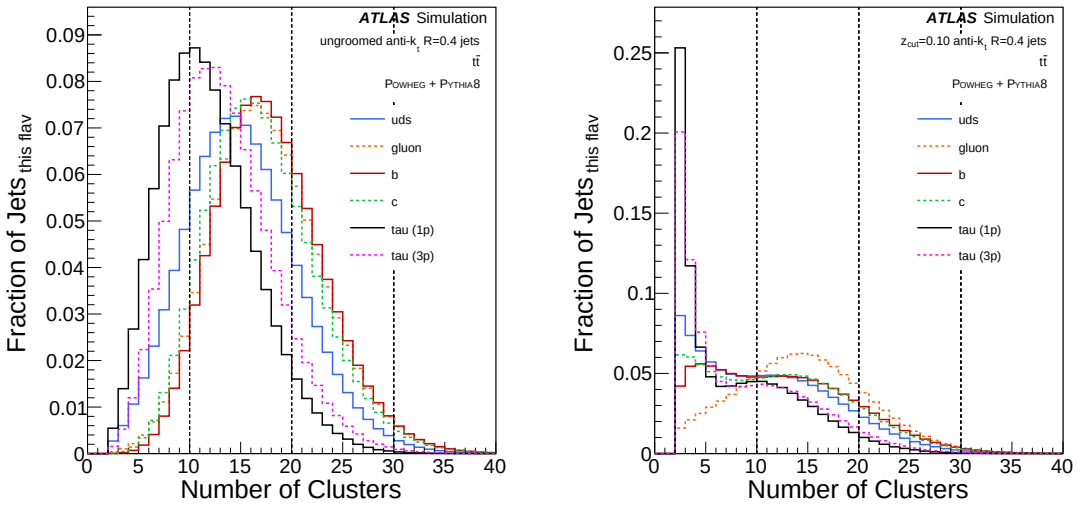
Groom name	β **value**	z **value**
SDb0z05	0	0.05
SDb0z1	0	0.10
SDb0z15	0	0.15
SDb0z20	0	0.20
SDb0z25	0	0.25
SDb0z30	0	0.30

Table 6.2: Grooming strengths used for soft drop grooming. Z_{cut} is a soft threshold, and β is an angular component. [23].

The process of applying soft drop grooming of different strengths can be illustrated by the effect on the number of clusters in the reco-jet, an example of this process for reco-jets that are truth-flavour –tagged as b -jets is shown in Figure 6.1.

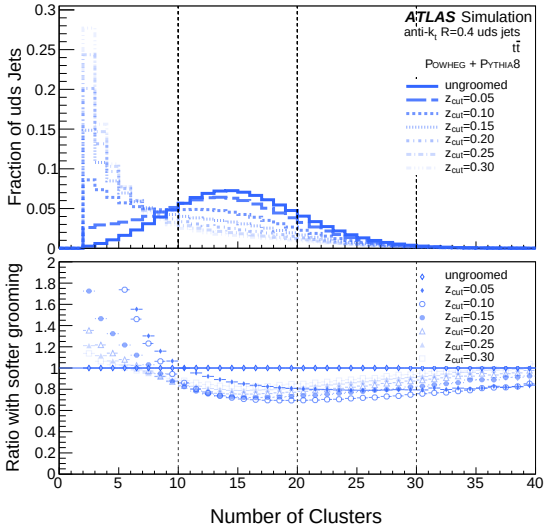
6.6 Soft Drop Grooming Flavour-Tagged Small Jets

6.6.1 Number of Clusters

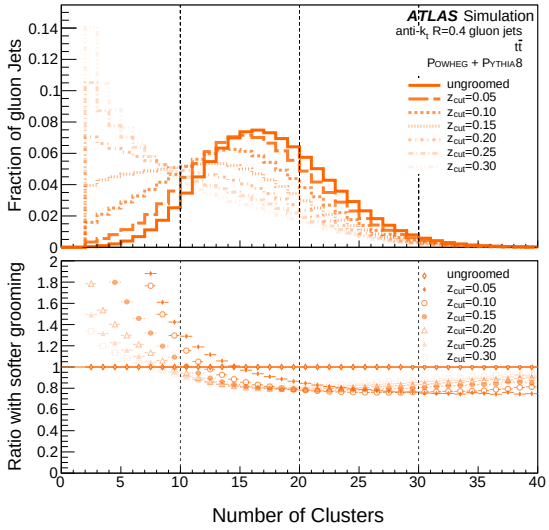


(a) ungroomed (b) $Z_{cut} = 0.1$

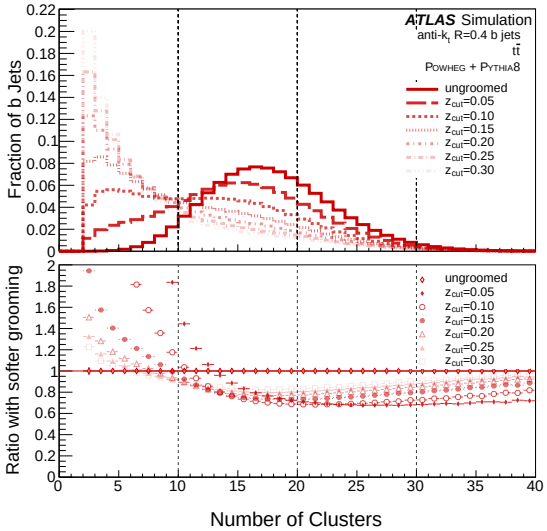
Figure 6.9: NCLUS for different truth-flavour tagged jets when ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b).



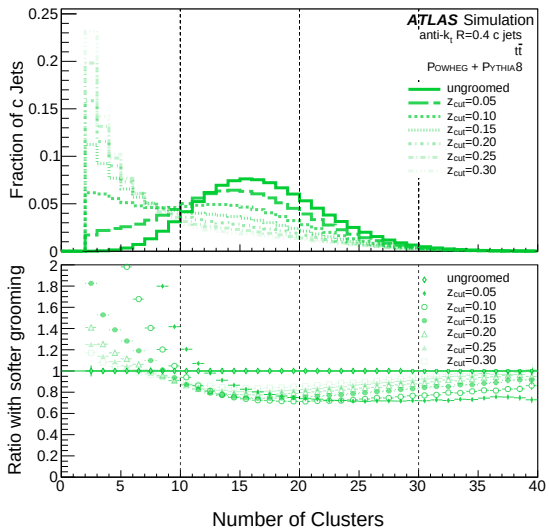
(a) uds



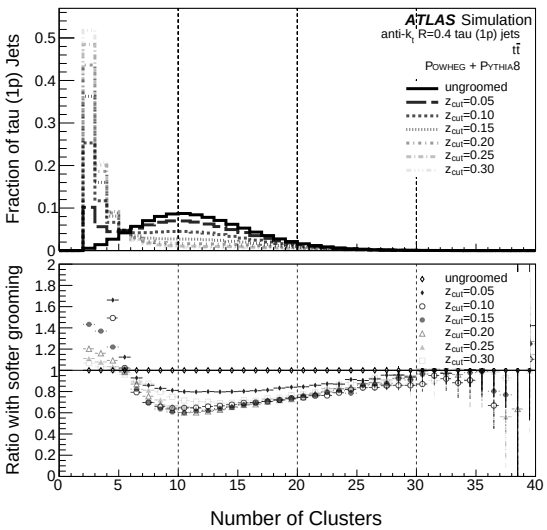
(b) gluon



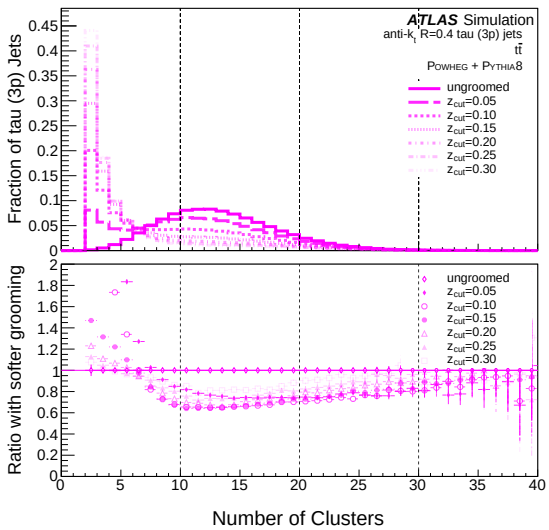
(c) *b*



(d) *c*



(e) 1-pronged tau



(f) 3-pronged tau

Figure 6.10: NCLUS at each strength of soft drop grooming with a ratio to the next softest grooming strength.

Throughout these analyses I focus on the effects of soft drop grooming on a number of observables, within both the reco-jets and truth-particle jets. The first of these is NCLUS, the number of topoclusters remaining in the reco-jet, both before and after various configurations of soft drop grooming are applied. The NCLUS distributions for ungroomed jets of different flavours are shown in Figure 6.9(a). What is shown is that tau-flavoured jets peak at a lower value of NCLUS than the other flavours, with 1-pronged tau - and 3-pronged tau -jets peaking at around ten and twelve clusters respectively. Following this, what is observed is uds-jets peaking at fourteen, c-jets peaking at 15, and both gluon-jets and b-jets peaking at 16 clusters. The width of each of these distributions are comparable.

The clearest indication of the effects of soft drop grooming comes from studying the number of calorimeter clusters remaining within the reconstructed jet at different levels of grooming. Firstly, the NCLUS found in the ungroomed flavour-tagged small jets is inspected.

Looking at Figure 6.9, the effects of soft drop grooming at the chosen grooming strength, $Z_{cut} = 0.10$, can be seen. The first thing to note is that tau-flavoured jets seem to be the most heavily groomed, with a very strong peak appearing at the minimum NCLUS, two. This is true more so for 1-pronged tau-jets rather than 3-pronged tau -jets. Secondly, very little change in the gluon-jets is noted: at this grooming strength they are the flavour of jet that changes the least. Jets sourced from quarks (uds, b, and c) lie somewhere between these two extremes, and at this grooming strength behave similarly.

What is expected is that jets with a well-defined hard-core, clear and distinct from those clusters indicative of wide-angle soft radiation, will lose more clusters following grooming. The significant imbalance of p_{T} between the clusters within the core and the clusters outside will lead to the low energy clusters being removed. The results shown here indicate, preliminarily, that the tau-flavoured jets have a well defined core by comparison. Those jets lacking the well-defined hard-core, such as gluon-jets, will therefore be less likely to lose clusters, and this is what is seen in the results.

The results for each of the grooming strengths, comparing flavours, can be seen in Figure A.1.

More detailed investigations into the effects of different grooming strengths upon each of these flavours was then undertaken. Figure 6.10 displays the NCLUS for each particular flavour of jet, at different strengths of soft drop grooming. The rate of change between one grooming strength and the next is also displayed, allowing identification of any thresholds with significant changes.

6.6.2 Sample Size

Another, related factor explored is the change in the number of flavour-tagged small jets of each flavour, and how this changes for each strength of soft drop grooming, shown in Figure 6.11.

A reminder that at least two clusters are required to remain within the jets to avoid rejection. If the jets have fewer than two clusters remaining following grooming then they are discarded.

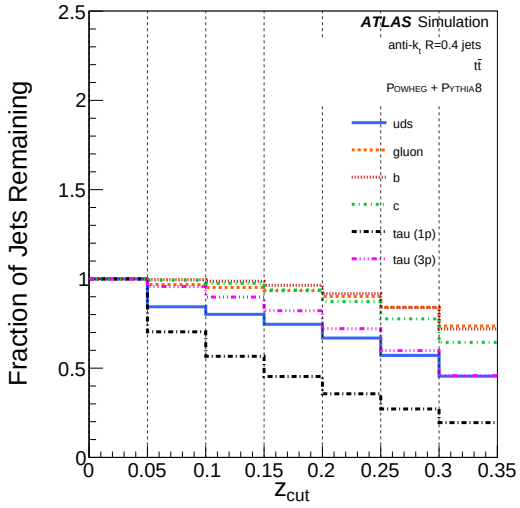


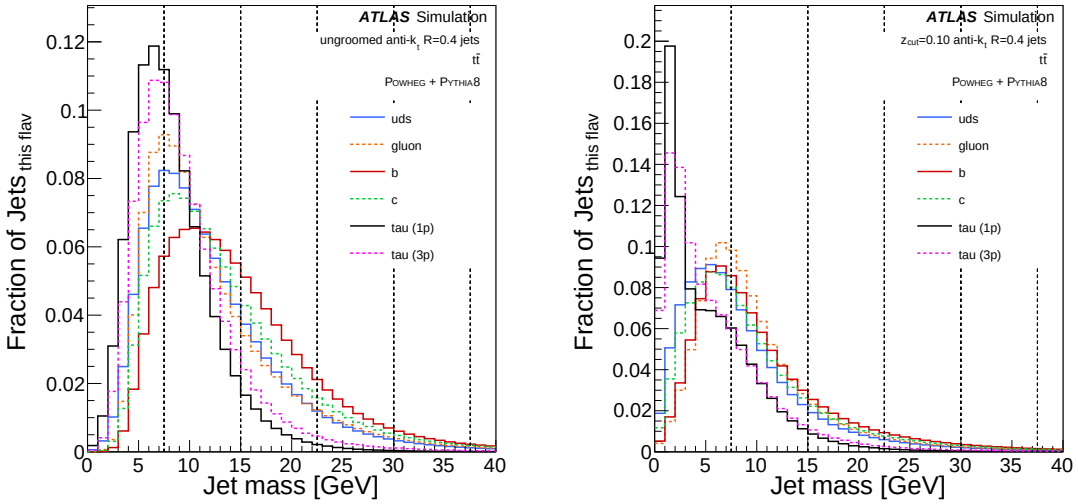
Figure 6.11: The fraction of each set of flavour-tagged small jets remaining following different strengths of soft drop grooming.

We see that overall, and at the first stage of grooming, 1-pronged tau–flavoured flavour-tagged small jets are most likely to be groomed away, quite dramatically compared to all other flavours: ultimately only 20% of these are left. The biggest change occurs between the ungroomed sample and that of the grooming strength $Z_{cut} = 0.05$, with subsequent increases in strength removing less and less of the original sample.

We see that the uds-flavour-tagged small jets and 3-pronged tau -flavoured jets are both heavily groomed away, though at different rates, losing close to 50% of the original sample size. As is the case with 1-pronged tau–flavoured flavour-tagged small jets, uds-jets lose most between the ungroomed sample and that of the grooming strength $Z_{cut} = 0.05$. Our *c*-, *b*- and gluon-jets lose relatively little in terms of sample size, only about 25% for gluons and *b*-jets. Both *c*- and *b*-jets do not lose any jets at all until grooming strength reaches $Z_{cut} = 0.15$. Our default grooming strength of $Z_{cut} = 0.10$ is the highest strength where all flavours of jets have more than 50% of their samples sizes left.

6.6.3 jet mass

The second JSS observable in this study is the jet mass of the reco-jet, both before and after soft drop grooming is applied. The jet mass distributions for ungroomed jets of different flavour are shown in Figure 6.12(a). Here, much the same order of peaks as in NCLUS is displayed, with 1-pronged tau - and 3-pronged tau -jetsf peaking at the lowest values of 6.54 GeV and 7.35 GeV, respectively. The only flavour to have moved in order is gluon-jets, now between tau- and quark-flavoured jets and peaking at 7.65 GeV. The flavours with the highest peak in jet mass are the quark-jets, with uds-, *c*- and *b*-jets peaking at 7.95 GeV, 8.55 GeV and 10.35 GeV respectively. Also note that the distribution for these values of jet mass are more varied than in NCLUS with tau-flavoured jets appearing more narrow than other flavours, and *b*-jets broader. This can be related to the results for NCLUS by looking at the ratio between the peak values for



(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.12: Jet mass for different truth-flavour tagged jets when ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b).

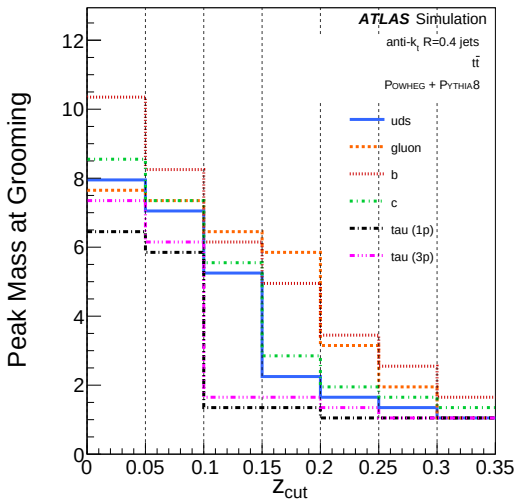


Figure 6.13: The peaks of the jet mass distributions for each set of flavour-tagged small jets at different strengths of soft drop grooming.

Flavour	Peak NCLUS value / Peak jet mass value
gluon	0.478
uds	0.568
c	0.570
3-pronged tau	0.613
1-pronged tau	0.645
b	0.647

Table 6.3: Ratio of peak value for NCLUS and jet mass in ungroomed jets of each flavour.

NCLUS and jet mass for each flavour, shown in Table 6.3 With gluon-jets there is a change in the order of peaks tied with the smallest ratio between the peak values of NCLUS and jet mass. This indicates that gluon-jets are generally composed of softer, lower-mass clusters. These results also indicate that tau- and *b*-jets have higher mass clusters, and are therefore more likely to have a hard-core of higher mass clusters.The order of this ratio from, lowest to highest, also corresponds to the mass of the isolated particles in ascending order.

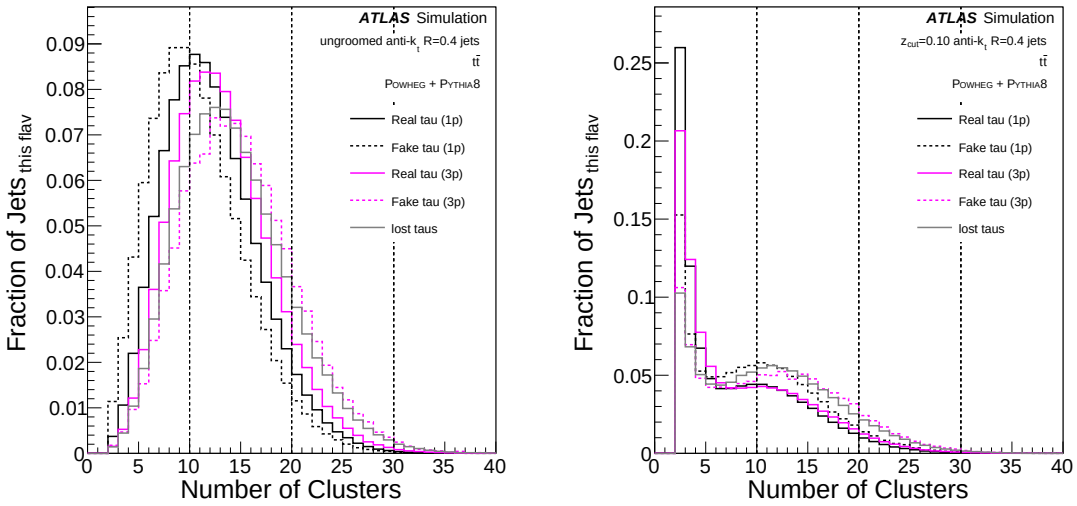
Next, the main JSS observable of interest, jet mass, was investigated. These results are shown in Figure 6.12. Here, some quite interesting results are seen following soft drop grooming at the chosen default strength of $Z_{cut} = 0.10$. Dramatic changes in the mass distribution of the tau-jets, particularly 1-pronged tau s is observed: the distributions spike quite dramatically compared to all the other flavour-tagged small jets, peaking at a value very close to that of the mass of the τ -lepton, 1.777 GeV. There also follows a second, much smaller peak. Our quark-flavour tagged jets show somewhat less serious changes, with uds only changing a little, *b*-jets changing by about half, and therefore the most and *c*-jets affected somewhere in between, with a shift of about a third. Following grooming the distributions of all three flavours are very similar. Also very interesting, the gluon-flavour-tagged small jets almost don't change at all at this level of grooming.

Comparing these changes with that of the NCLUS, a number of things are observed. Firstly the small change in gluon mass distribution is unsurprising as very little change in the NCLUS was seen. The changes in the tau-jets indicate that the remaining mass, close to that of the τ -lepton, is held in very few clusters, just 2 or 3, this may be indicative of a hard-core. However, significant jet mass lay outside of these clusters too. It is harder to come to any definitive conclusions about the changes seen in the quark jets, but the relation seems to indicate that each of the clusters that were removed potentially held a smaller portion of the mass of those that remain, as the mass changes less than the NCLUS. The results for each of the grooming strengths, comparing flavours, can be seen in Figure A.2.

To investigate this shift in more detail, the peak of the mass distributions at different strengths of soft drop grooming were compared, shown in Figure 6.13. Here, another interesting result for the tau-jets was seen, a dramatic drop in mass between the soft drop grooming $Z_{cut} = 0.05$ and $Z_{cut} = 0.10$, much higher than the difference at any other level of grooming. We see the biggest

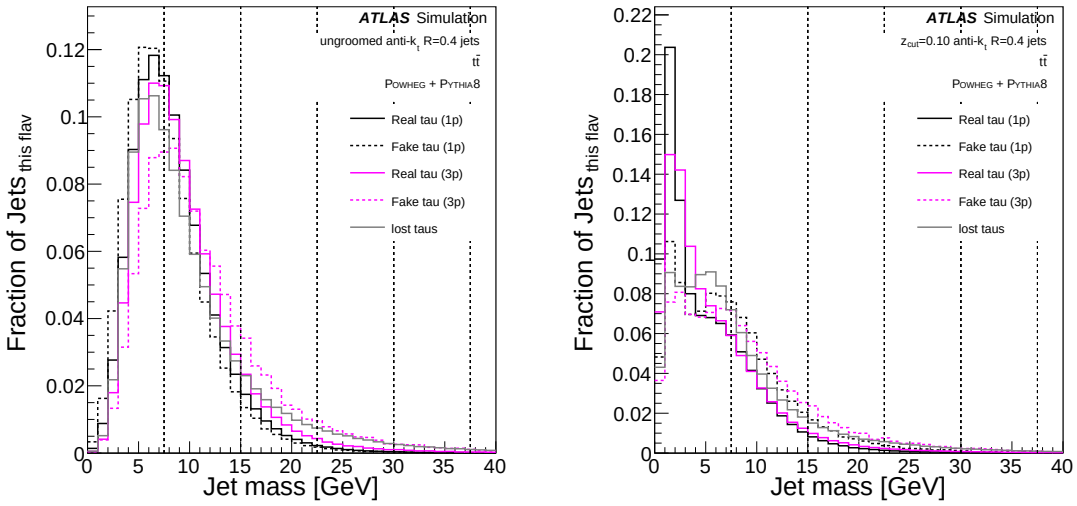
drop in uds-jets and c-jets taking place at the next grooming strength, $Z_{cut} = 0.15$, gluons at $Z_{cut} = 0.20$ and b-jets showing a steady decrease at each step.

6.7 Soft Drop Grooming with Tau-Jets



(a) ungroomed (b) $Z_{cut} = 0.10$

Figure 6.14: NCLUS for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), tau flavour-tagged jets.



(a) ungroomed (b) $Z_{cut} = 0.10$

Figure 6.15: Jet mass for the ungroomed, (a), and soft drop groomed, (b), tau flavour-tagged jets.

As seen from Figures 6.9 and 6.12, the effects that soft drop grooming has on the tau-jets is very interesting. To investigate these effects in greater depth, two areas were explored:

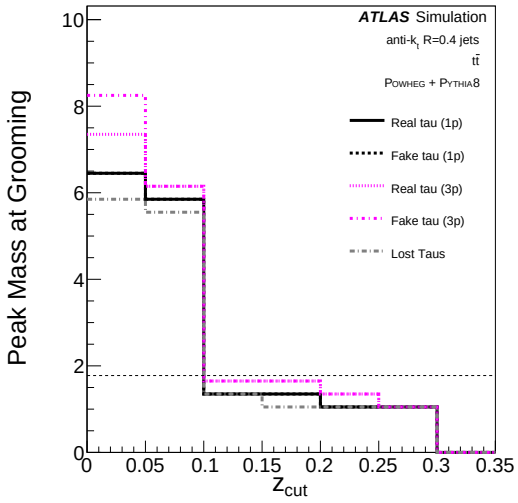


Figure 6.16: The peaks of the jet mass distributions for each set of tau-jets at different strengths of soft drop grooming.

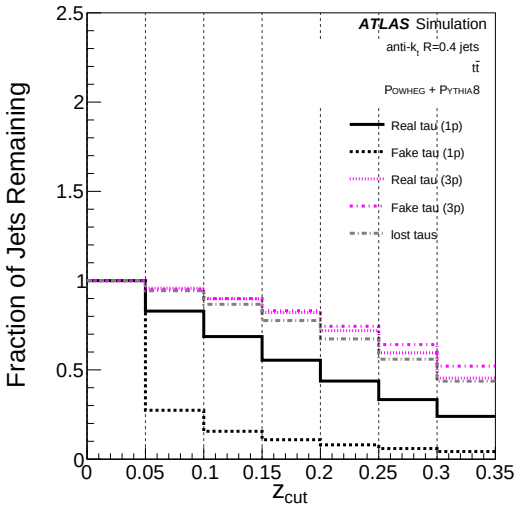


Figure 6.17: The fraction of each set of tau-jets remaining following different strengths of soft drop grooming.

firstly, whether soft drop grooming allows discrimination between incorrectly identified tau-jets (fakes) and correctly identified tau-jets; secondly, whether tau-jets, which have not been identified using traditional τ-tagging methods, can be correctly identified. For this purpose the properties of five sub-categories of tau-jet were investigated (see Section 6.4.2 and Table6.4).

We can start by looking at the effects of grooming on the mass of and NCLUS in each of tau-jets types. In Figure 6.14 the NCLUS distribution for tau-jets, both before and after applying soft drop grooming with strength $Z_{cut} = 0.10$, is presented. As previously observed, following grooming the NCLUS strongly peaks at the minimum value of two clusters. In Figure 6.15 the effects of soft drop grooming on the mass of the jets are displayed. It can be observed that the pronounced peak at low mass is present in all flavours of tau-jets , though to various degrees.

Tau Jet Type	Description
Real 1-pronged tau	Matched to a traditional tau jet that has one ID track and is matched to a truth-particle tau.
Real 3-pronged tau	Matched to a traditional tau jet that has three ID track and is matched to a truth-particle tau.
Fake 1-pronged tau	Matched to a traditional tau jet that has one ID track and is nott matched to a truth-particle tau.
Fake 3-pronged tau	Matched to a traditional tau jet that has three ID track and is not matched to a truth-particle tau.
Lost taus	Not match to a traditional tau jet.

Table 6.4: The types of tau jet used in this analysis.

Looking also at the peak jet mass for tau-jets at each grooming strength some common behaviour can be observed. In Figure 6.16 it is apparent that applying grooming increasingly lowers the peak jet mass of tau-jets , but all tau-jets have a severe drop in peak jet mass at soft drop grooming strength $Z_{cut} = 0.10$, dropping just below the τ -lepton mass of 1.777 GeV.

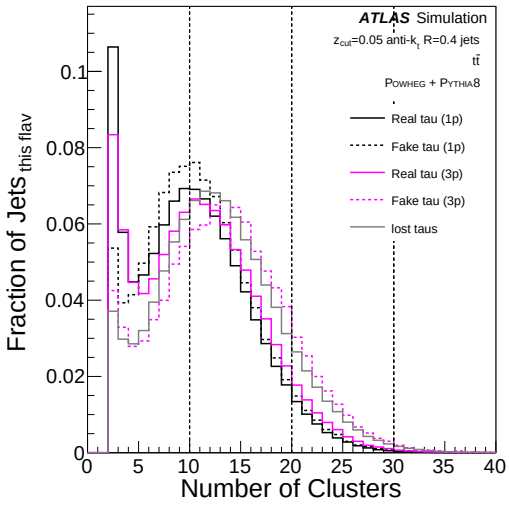
The effects of soft drop grooming on the number of jets in the samples can also be investigated. In Figure 6.17 it can be seen that (for most tau-jets) applying soft drop grooming steadily removes a portion of the sample. A noticeable difference between 1-pronged tau-jets and 3-pronged tau -jets can also be seen, with 1-pronged tau depleting more rapidly than 3-pronged tau . The single-pronged nature of 1-pronged tau likely leads to a higher portion of jets groomed down to a single cluster, and thus no longer passing the minimum requirement of two clusters. Most striking, however, is the difference between fake and real 1-pronged tau-jets. See Section 6.7.2.

6.7.1 Recovering Lost Tau-Jets

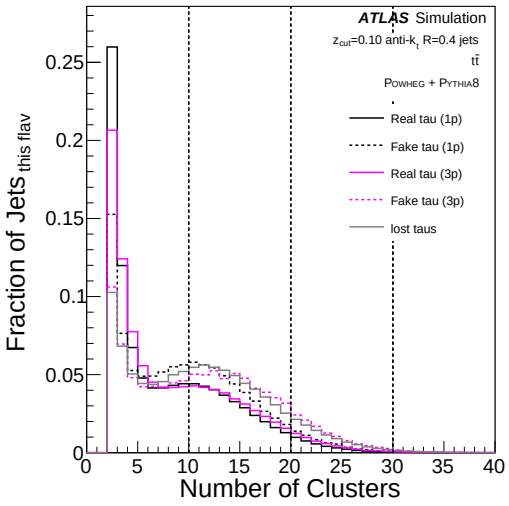
Firstly, investigations into whether lost tau-jets can be potentialy identified from groomed observables is undertaken. Looking at Figure 6.14, the effects of soft drop grooming on the distribution of NCLUS within tau-jets can be seen. As previously observed, following grooming the NCLUS strongly peaks at the minimum value of two clusters. This is true for all tau-jets including lost tau-jets, although initially less pronounced. Looking at the distribution of NCLUS within tau-jets for all groomings, in Figure 6.18, it is apparent that this is a trend that continues with increased grooming strength.

Investigating the effects of soft drop grooming on the jet mass of lost tau-flavoured jets again demonstrates that the trend common to all tau-jets is present: a peak present between 1 GeV and 2 GeV. This peak is much less prominent for lost tau-jets, but the start of it is there. Looking at Figure B.1 it is shown that this peak becomes more apparent at higher strengths of grooming. These lost taus-jets can also be compared to non-tau flavours in Figure B.9 and Figure B.10.

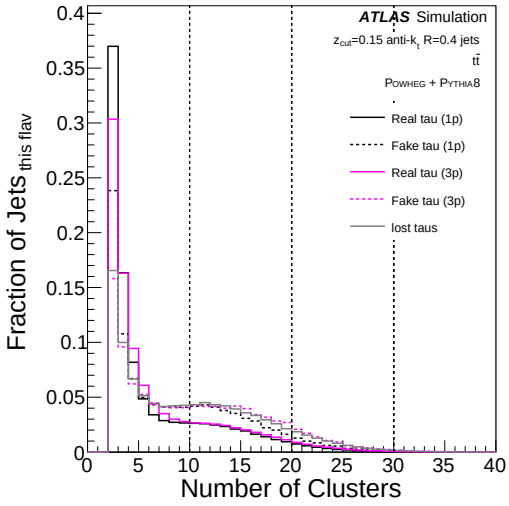
Using soft drop grooming on these reveals a common behaviour with the other tau-jets, with early peaks appearing, both in NCLUS and jet mass, and growing in prominence at higher



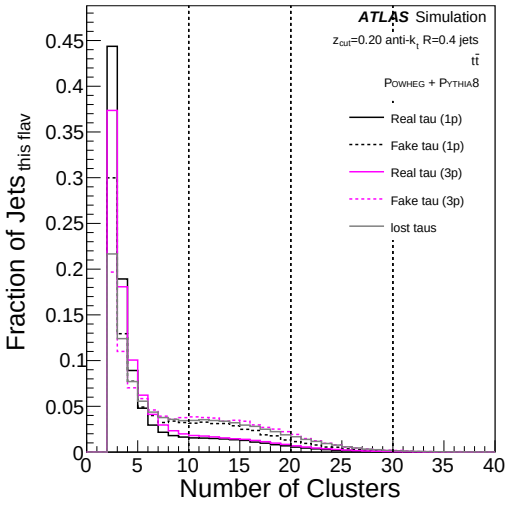
(a)



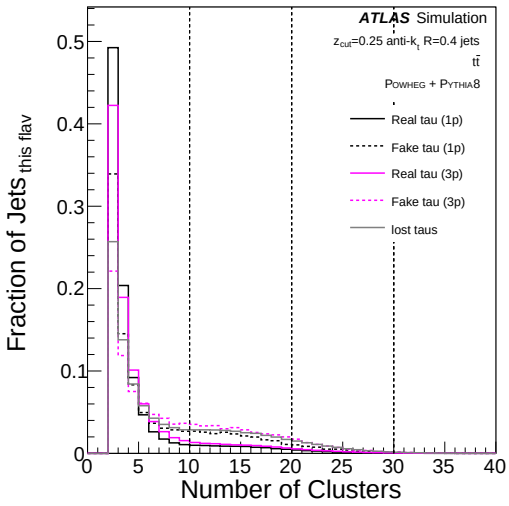
(b)



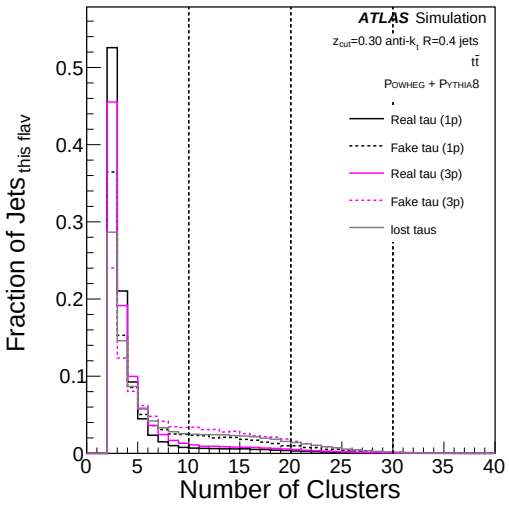
(c)



(d)



(e)

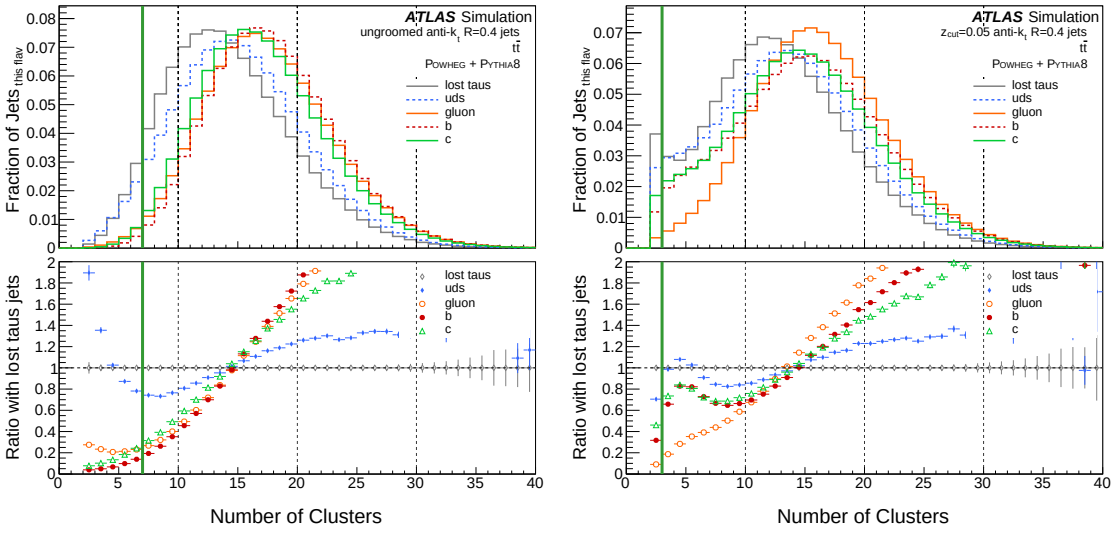


(f)

Figure 6.18: Number of clusters for tau-flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

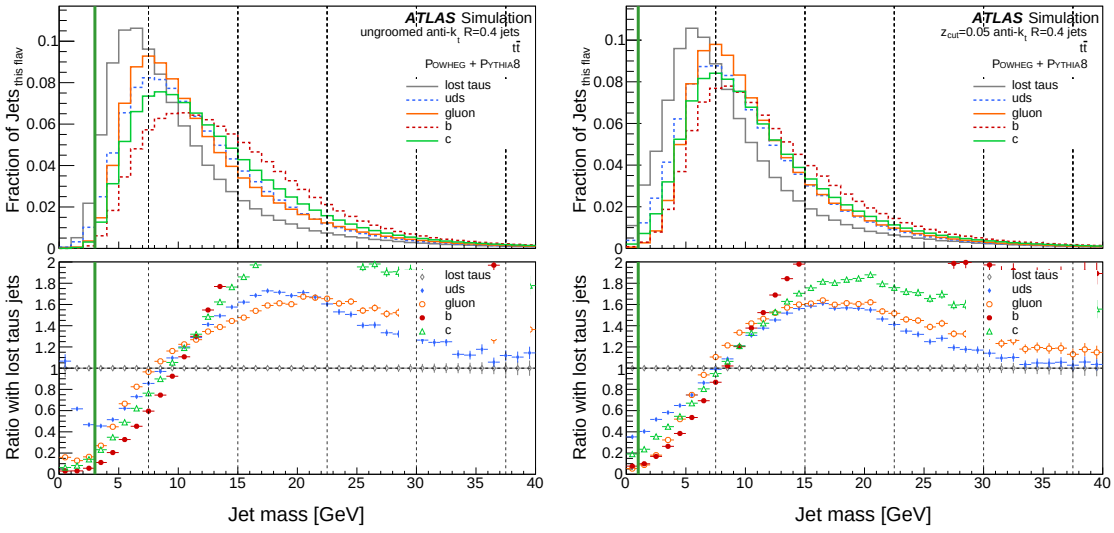
strengths of grooming. Explorations, therefore, were undertaken to test if it is possible to introduce possible selection cuts, relying on these features, to separate lost tau-jets from other, non-tau-jets.

Selection Cuts on lost tau-jets



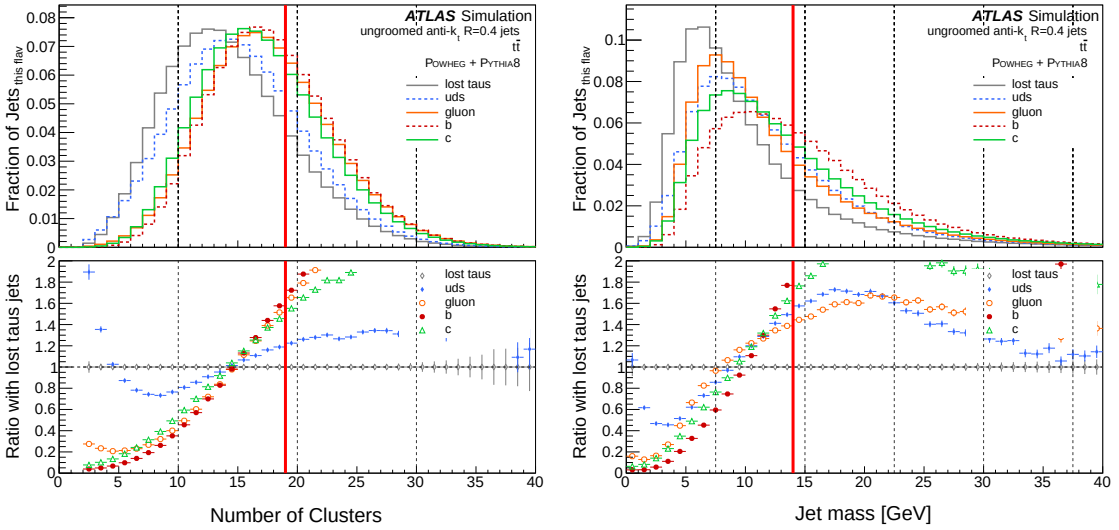
(a) ungroomed (b) $Z_{cut} = 0.05$

Figure 6.19: NCLUS in lost tau-jets and all non-tau-jets ungroomed, (a), and with soft drop grooming strength $Z_{cut} = 0.05$, (b), with a ratio to the number of lost tau-jets



(a) ungroomed (b) $Z_{cut} = 0.05$

Figure 6.20: Jet mass in lost tau-jets and all non-tau-jets ungroomed, (a), and with soft drop grooming strength $Z_{cut} = 0.05$, (b), with a ratio to the number of lost tau-jets. The green line indicates the selection cut.



(a) NCLUS (b) jet mass

Figure 6.21: NCLUS, (a), and jet mass, (b), in ungroomed lost tau-jets and all non-tau-jets, with a ratio to the number of lost tau-jets. The red line indicates the selection cut with an 80 % WP.

Using soft drop grooming has a small effect on improving the identification and selection of lost tau-jets through the use of selection cuts. How soft drop grooming improves selection cuts on NCLUS can be explored. At ungroomed level the best possibly purity that can attainain from the sample is by only selecting jets with fewer than seven clusters. Of the remaining jets, 1.71 % are lost tau-jets. This retains 6.44 % of all lost tau-jets in the sample. This purity can only be slightly improved upon through the use of soft drop grooming, but with a significant loss of the sample size. At soft drop grooming strength $Z_{cut} = 0.05$ the best purity possibly purity that can be attained from the sample is by only selecting jets with fewer than three clusters. This gives a purity of 1.73 %, but retains only 3.7 % of soft drop grooming lost tau-jets. Both of these potential cuts are shown Figure 6.19.

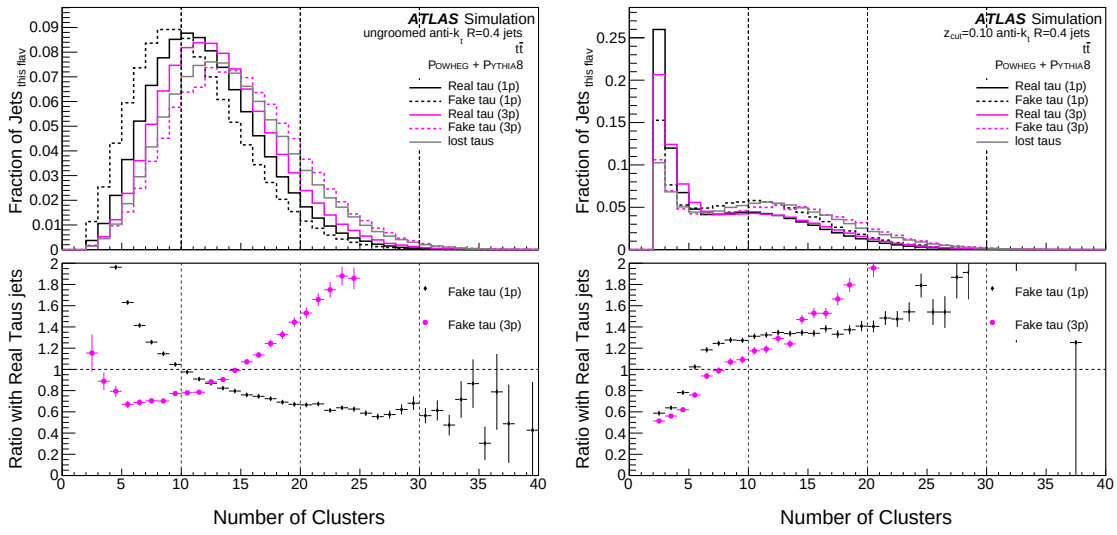
Looking also at possible selection cuts on jet mass there is a similar trend. These are shown in Figure 6.20. At ungroomed level, a maximum possible purity of 2.95 % is achievable, by only selecting jets with a jet mass less than 3 GeV. This keeps 2.75 % lost tau-jets. It is possible to improve upon this purity with soft drop grooming jets with a strength of $Z_{cut} = 0.05$ and selecting only those remaining jets with jet mass less than 1 GeV. The purity increases to 4.12 %, but with the remaining sample size dropping to 1.03 %.

Choosing from only cuts that return 80 % of lost tau-jets at the given strength of soft drop grooming, it is found that using grooming reduces the purity of the samples. First, this selection cut is applied on NCLUS. The single cut of this type, retaining at least 80 % of the groomed sample, with the best given purity has me only keep jets with fewer than nineteen clusters. This cut retains 81.5 % of lost tau-jets, which make up 0.809 % of all remaining jets. This purity cannot be improved upon through the use of soft drop grooming and selection cuts on NCLUS whilst retaining at least 80 % of jets.

Following the same process for jet mass the same result is observed, the highest purity attainable through the use of selection cuts whilst retaining at least 80 % of jets is at ungroomed level. Selecting only jets with jet mass less than 14 GeV, a purity of 0.842 % is obtained, with 81.6 % of lost tau-jets remaining from the original sample. The cuts for the chosen selection are shown in Figure 6.21.

These results initially suggest that identifying and distinguishinglost tau-jets from all the other flavours is not obviously made easier just through the use of soft drop grooming, although the small number of lost tau-jets, in relation to non-tau-jets , has a strong part to play in this. The observable features common to tau-jets following soft drop grooming may still be useful a tool for recovering these lost tau-jets.

6.7.2 Identifying Fake Tau-Jets

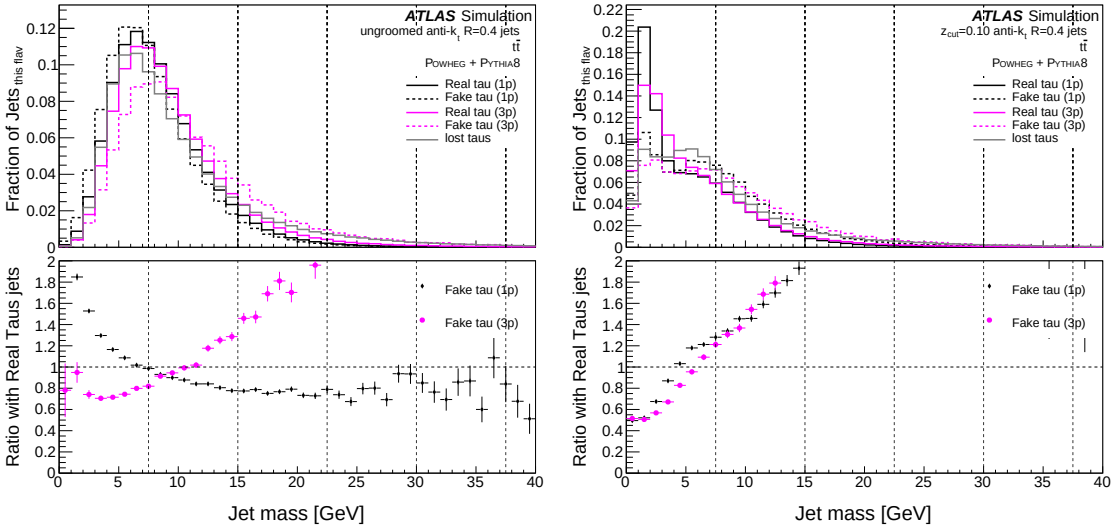


(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.22: NCLUS for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), tau flavour-tagged jets.

We can explore the differences between “real” and “fake” tau-jets within each bin in more detail by including a ratio plot displaying the ratio between these “fake” and “real” tau-jets , showing how they diverge. We do this for both 1-pronged tau and 3-pronged tau . We display these for ungroomed jets in Figure 6.22 and soft drop grooming with defaul strength, $Z_{cut} = 0.10$ in Figure 6.23. For the NCLUS within the jets, this is shown, for all grooming strengths in Figure B.13. For jet mass this is shown, for all grooming strengths, in Figure B.14.

As previously observed, following grooming the NCLUS strongly peaks at the minimum value of two clusters. This is true for all tau-jets , but is most prominent in the real tau-jets . The distributions for jet mass show peaks at the τ -lepton mass, much more prominently for real than for fake tau-jets . We see, in both cases, soft drop grooming acting more strongly on real tau-jets than fakes. Using soft drop grooming on the tau-jets reveals a divergence in behaviour



(a) ungroomed (b) $Z_{cut} = 0.10$

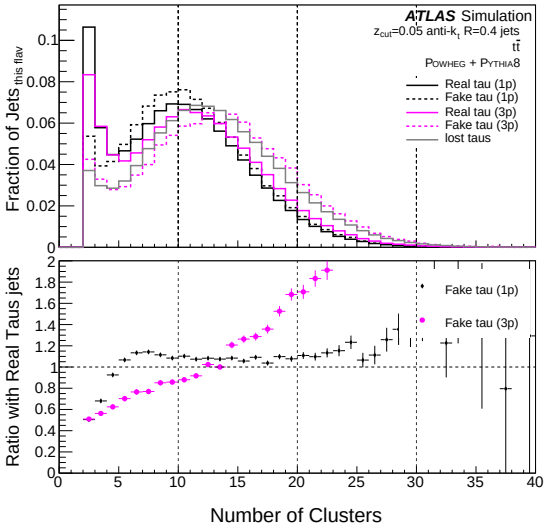
Figure 6.23: Jet mass for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), tau flavour-tagged jets.

between real and fake tau-jets . The distinctive tau-jets behaviour of early peaks appearing, for both NCLUS and jet mass, and growing in prominence at higher strengths of grooming, is less pronounced for fake tau-jets . We can therefore, explore possible selection cuts, utilising these features, to separate the real tau-jets from fakes. We can also see from Figure 6.17 that there are obvious differences in the remaining sample sizes of real and fake 1-pronged tau-jets following soft drop grooming strength of $Z_{cut} = 0.05$, with the fake 1-pronged tau-jets dropping close to four times as much as those that are real.

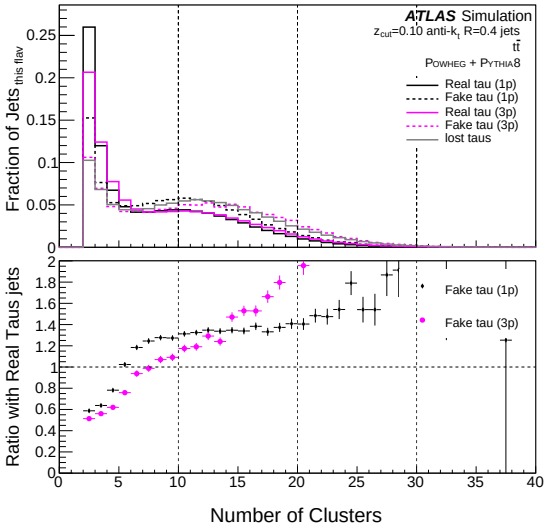
Looking at Figure 6.17 the differences in sample-loss between fake and real taus can be observed. Close to 75 % of fake 1-pronged tau-jets are lost following soft drop grooming of strength $Z_{cut} = 0.05$, with the effects of grooming waning at higher strengths, compared to less than 20 % of real 1-pronged tau-jets. This indicates these fake 1-pronged tau-jets have a single-cluster hard-core with much softer surrounding clusters. The reverse behaviour for 3-pronged tau -jets, whereby fake 3-pronged tau -jets are less likely to be groomed away, indicates a less pronounced hard-core than in those that are real.

Selection Cuts on fake tau-jets

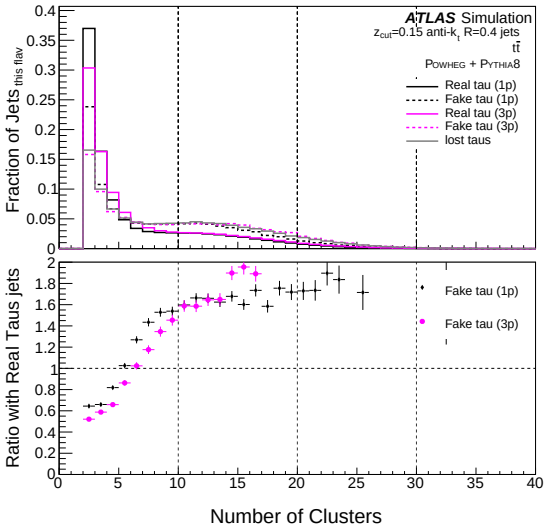
Using soft drop grooming on the tau-jets seems promising for improving the identification and separation of real tau-jets from fake through the use of selection cuts. We first look at the effect of soft drop grooming on using selection cuts on the NCLUS distribution to separate real and fake tau-jets . At ungroomed level the best cut available removes all jets with fewer than 22 clusters. Of the tau-jets found remaining, 86.3 % are real. However, this cut leaves only 3.07 % of -pronged tau-jets and 1.88 % of 3-pronged tau . This is reasonably improved upon through the use of soft drop grooming. A jump in possible sample purity at a strength of $Z_{cut} = 0.05$, to



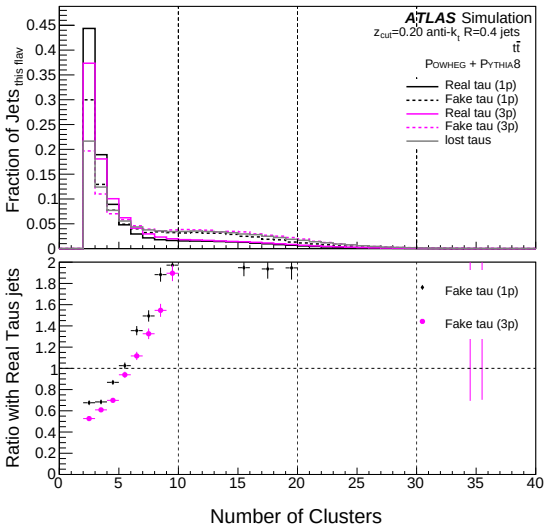
(a)



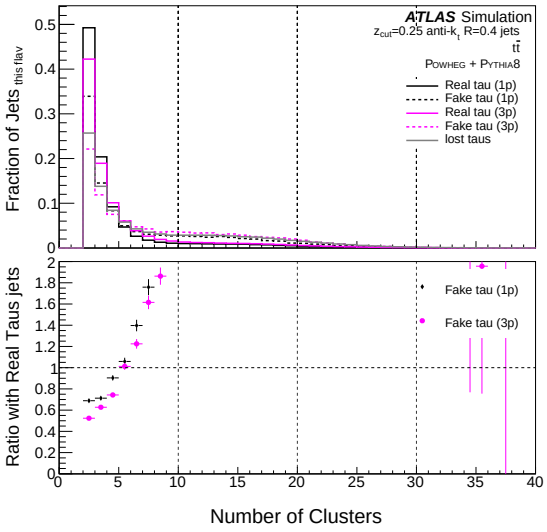
(b)



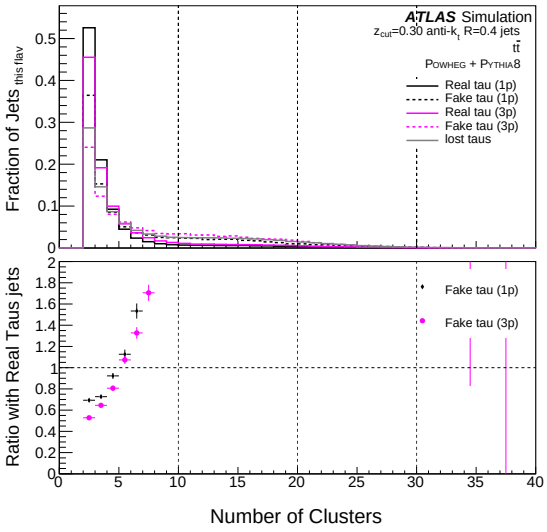
(c)



(d)



(e)



(f)

Figure 6.24: Number of clusters for tau-flavour jets, groomed with strength Z_{cut} = 0.05, (a), Z_{cut} = 0.10, (b), Z_{cut} = 0.15, (c), Z_{cut} = 0.20, (d), Z_{cut} = 0.25, (e), Z_{cut} = 0.30, (f), all normalised to the number of real taus for each prong value.

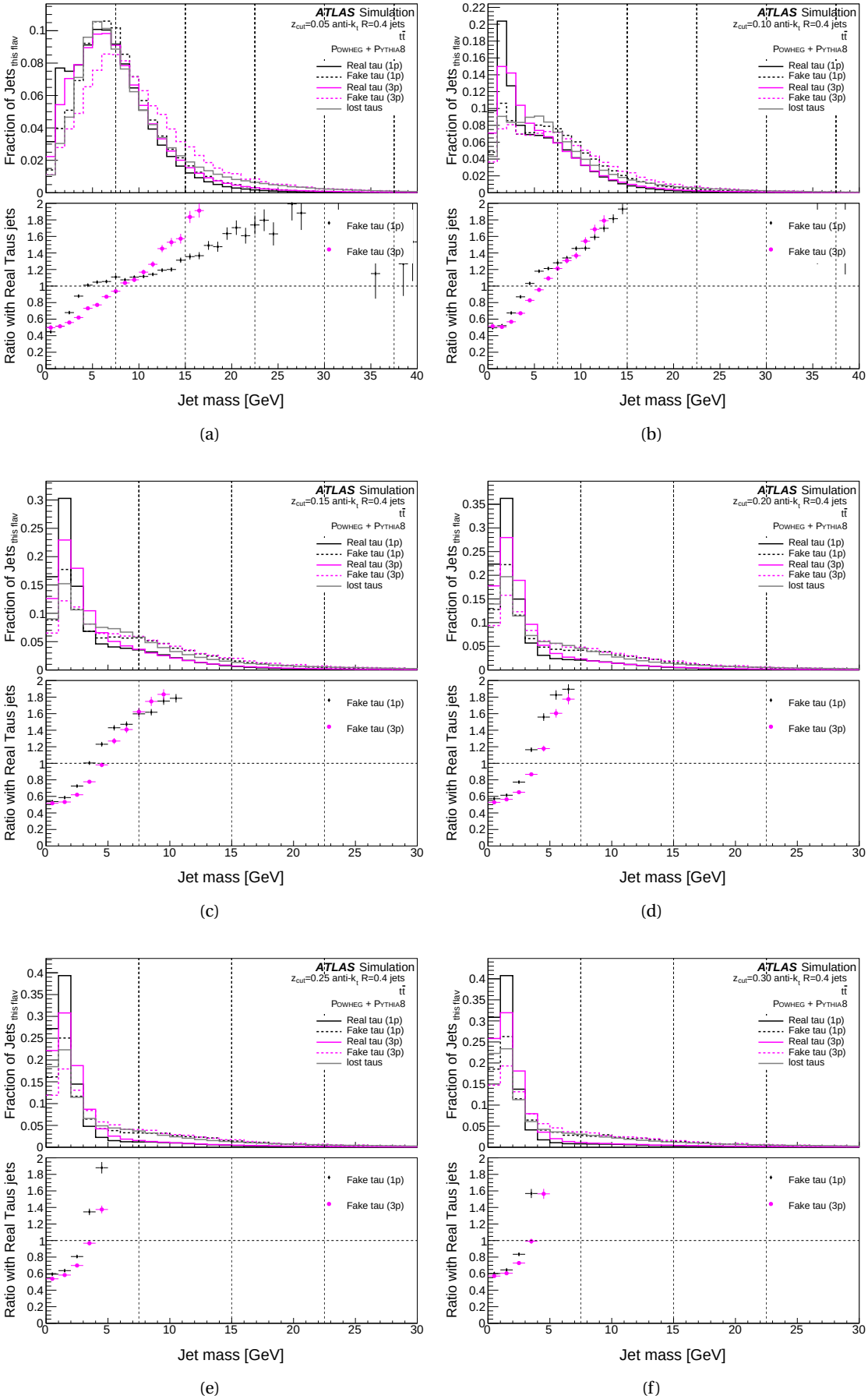
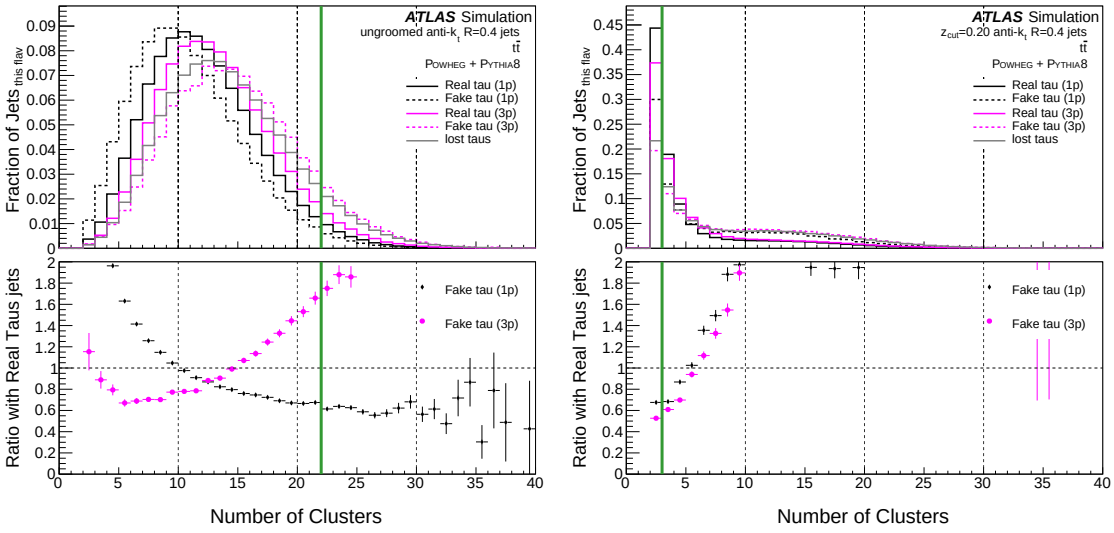
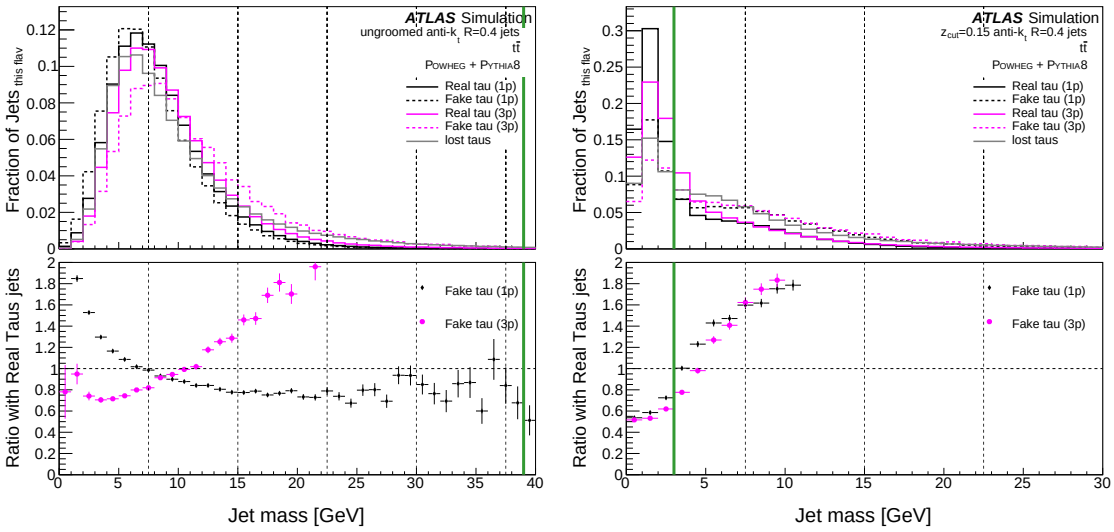


Figure 6.25: Jet Mass for tau-flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of real taus for each prong value.



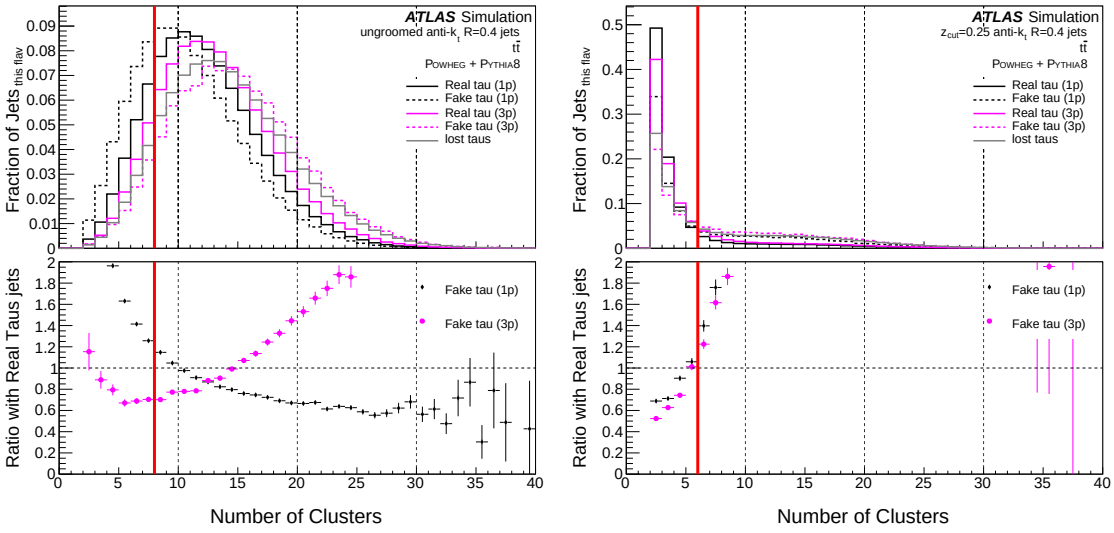
(a) ungroomed (b) $Z_{cut} = 0.2$

Figure 6.26: NCLUS in lost tau-jets and all non-tau-jets ungroomed, (a), and with soft drop grooming strength $Z_{cut} = 0.2$, (b), with a ratio to the number of real tau-jet. The green line indicates the selection cut.



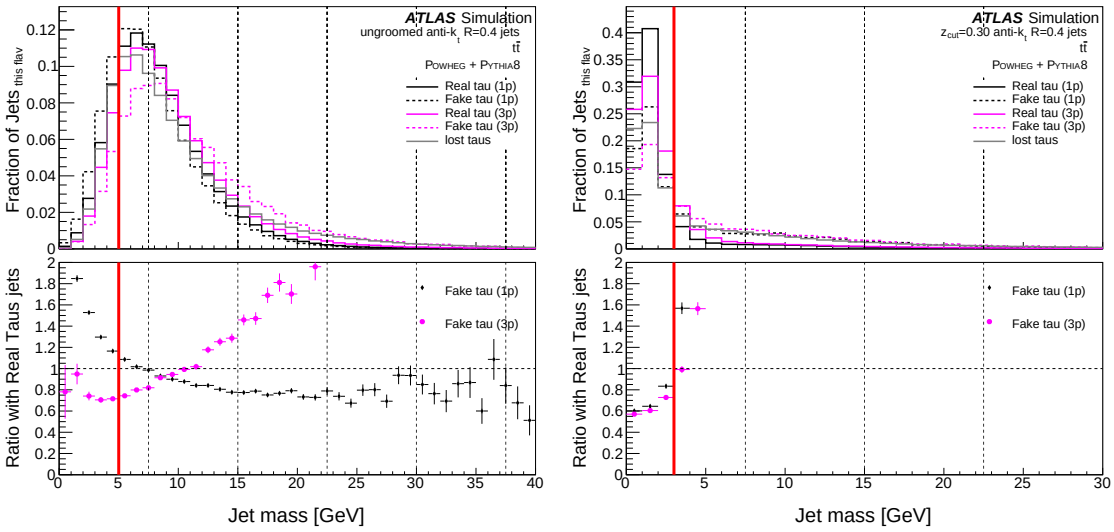
(a) ungroomed (b) $Z_{cut} = 0.15$

Figure 6.27: Jet mass in lost tau-jets and all non-tau-jets ungroomed, (a), and with soft drop grooming strength $Z_{cut} = 0.15$, (b), with a ratio to the number of real tau-jet. The green line indicates the selection cut.



(a) ungroomed (b) $Z_{cut} = 0.25$

Figure 6.28: NCLUS in lost tau-jets and all non-tau-jets ungroomed, (a), and with soft drop grooming strength $Z_{cut} = 0.25$, (b), with a ratio to the number of real tau-jet. The red line indicates the selection cut with an 80 % WP.



(a) ungroomed (b) $Z_{cut} = 0.3$

Figure 6.29: Jet mass in lost tau-jets and all non-tau-jets ungroomed, (a), and with soft drop grooming strength $Z_{cut} = 0.3$, (b), with a ratio to the number of real tau-jet. The red line indicates the selection cut with an 80 % WP.

95.8 %, coincides with the significant drop in fake 1-pronged tau-jets. By applying a soft drop grooming strength of $Z_{cut} = 0.20$ and selecting only those tau-jets with fewer than three clusters a purity of 96.6 % is achieved. This cut also retains 44.4 % of 1-pronged tau-jets and 30.0 % of 3-pronged tau -jets, a significant increase. These cuts can be seen in Figure 6.26.

Moving to jet mass a similar development can be seen: at ungroomed level the highest purity that can be obtained is 86.7 %, by removing all tau-jets with a jet mass lower than 39 GeV. Following this selection cut, only 0.008 95 % of 1-pronged tau-jets and 0.004 59 % 3-pronged tau -jets remain. We again see a significant jump in the highest possible purity following the application of soft drop grooming with a strength of $Z_{cut} = 0.05$, to 96.1 %. At $Z_{cut} = 0.15$ it is possible to achieve a purity of 96.7 % by selecting only those tau-jets with a jet mass of less than 3 GeV. This keeps 61.5 % of 1-pronged tau-jets and 37.3 % of 1-pronged tau3-jets. Once more, a dramatic gain in the number of jets in the high-purity region. These cuts can be seen in Figure 6.27.

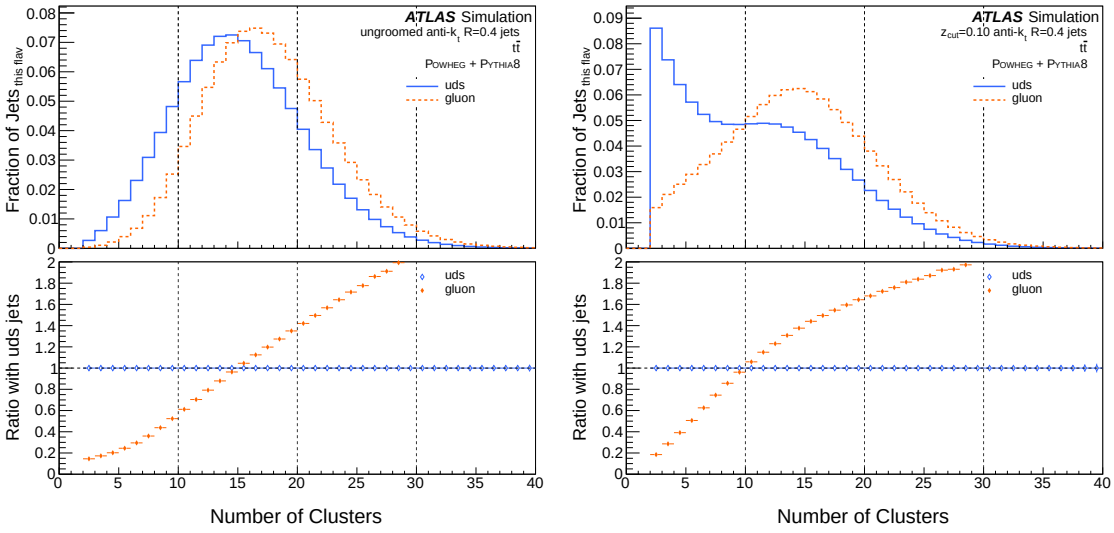
Moving onto selection cuts that satisfy the 80 % WP, improvements in the purity attainable from the tau-jets can be seen. For the ungroomed tau-jets, wash is found is that a selection that removes any jets with fewer than 8 clusters can achieve a purity of 83.4 %, retaining 80.9 % of 1-pronged tau-jets and 87.1 % 3-pronged tau -jets. Following soft drop grooming of strength $Z_{cut} = 0.05$ this increases to 92.4 %. The highest value acquired is 96.2 %, by selecting tau-jetsgroomed with a strength of $Z_{cut} = 0.25$ and with less than 6 clusters. After this cut has been applied 83.5 % of 1-pronged tau and 87.1 % 3-pronged tau -jets remain. These cuts can be seen in Figure 6.28.

For jet mass, the behaviour is the same. The purity of ungroomed jets can be maximised, to a value of 82.4 %, by removing all tau-jetswith jet mass less than 5 GeV. This selection keeps 81.4 % of 1-pronged tau-jets and 85.8 % of 3-pronged tau -jets. At soft drop grooming strength $Z_{cut} = 0.05$ this increases to 92.4 %. We reach a maximum at soft drop grooming strength $Z_{cut} = 0.30$, by cutting to keep only those tau-jets with a jet mass of less than 3 GeV. From this selection a purity of 96.4 % is reached, keeping 85.4 % of 1-pronged tau-jets and 75.9 % of 3-pronged tau -jets. These cuts can be seen in Figure 6.29.

In each of these cases soft drop grooming provides a definitive improvement in the ability to separate real and fake taus, with higher purities available, and less reduction in sample size occurring from applying selection cuts. Ultimately, the effect soft drop grooming has on removing a large portion of fake 1-pronged tau-jets seems to play a significant part in improving this separation.

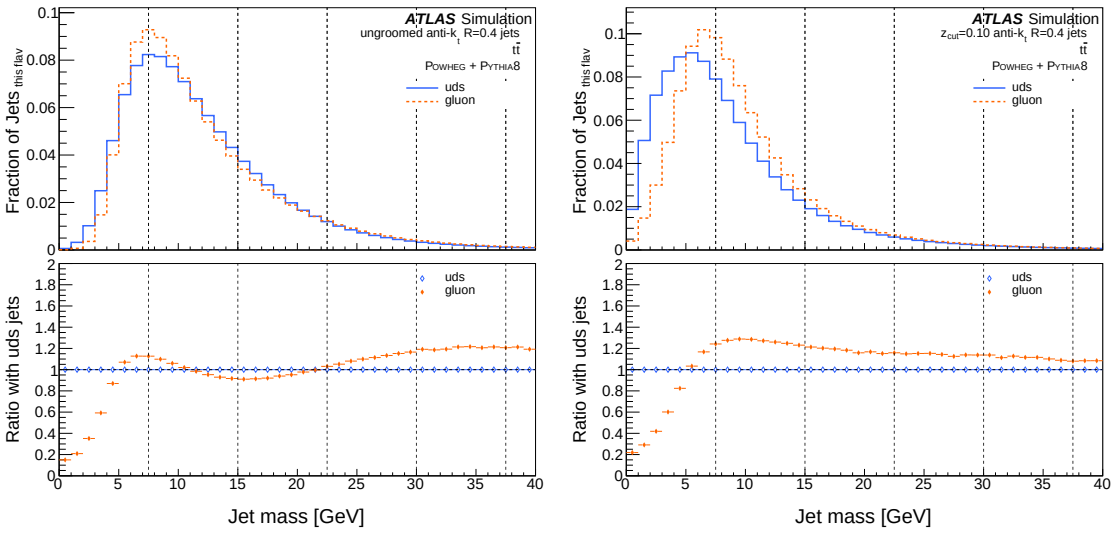
6.8 Light jets

The effects that soft drop grooming has on gluon initiated jets, leads to investigations in more detail the effects on grooming on both flavours of light jets: uds and gluons. Could soft drop grooming be used to help discriminate between uds- and gluon-flavour-tagged small jets? Here comparisons between these two flavours is presented. We display the number of jets as a rel-



(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.30: NCLUS for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), light flavour-tagged jets, normalised to the number of uds-jets with a ratio to uds-jets.



(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.31: Jet mass for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), light flavour-tagged jets, normalised to the number of uds-jets with a ratio to uds-jets.

ative portion of uds-jets. We also display a ratio plot beneath, to compare gluon to uds within each bin in closer detail.

We can start by looking at the effects of grooming on the NCLUS and jet mass for both uds- and gluon-jets. In Figure 6.30 the NCLUS distribution in these jets before and after applying soft drop grooming with strength $Z_{cut} = 0.10$ is displayed.

As noted previously in Figure 6.9 the change in the distribution of uds-jets leads to a high peak at the minimum value for NCLUS, two. The gluon-jet distribution, however, are changed very little, but has broadened and shifted to smaller values.

Similarly, looking at the effect of soft drop grooming on jet mass, shown in Figure 6.31, only a very small shift in the distribution of gluon-jets, to smaller values, is seen. The change in uds-jets this grooming strength is not dramatically different, however more prominent differences start to appear at higher strengths of soft drop grooming.

NCLUS at all grooming strengths, is shown in Figure C.3. For jet mass this is shown, for all grooming strengths, in Figure C.4.

As discussed in Section 6.6.3 differences in the sample-loss of light-jets following soft drop grooming is observed, with close to twice as many uds-jets than gluon-jets removed. Section 6.6.3 also shows the different changes in peak-jet mass in light-jets after application of soft drop grooming, with the peak-jet mass of uds-jets lower than quark-jets for many groomings.

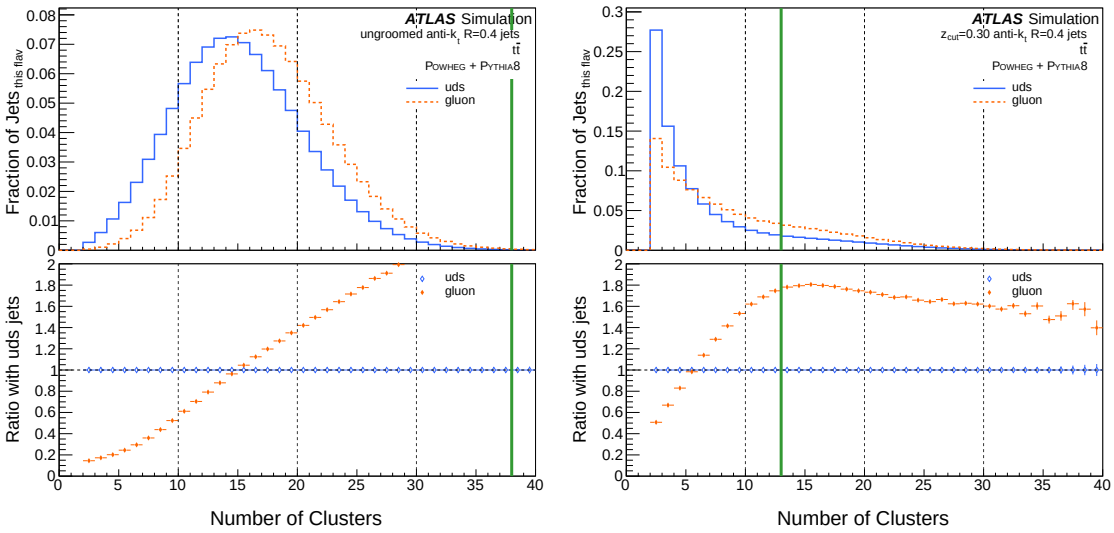
The differences in the ways uds- and gluon-jets behave under soft drop grooming allow explorations of whether groomed jets can be utilised to better separate these flavours through the use of selection cuts.

Selection Cuts on light-jets

Investigations as to whether application of soft drop grooming can bring improvement of the process of making selection cuts to maximise the purity of the sample, either for uds-jets for gluon-jets was undertaken. First, investigations were undertaken into whether any cuts on the NCLUS within soft drop grooming jets can improve signal purity when looking for either uds- or gluon-jets amongst the light-flavoured jets. For uds-jets this does not seem to be the case, the highest obtainable purity is amongst ungroomed jets, where 89.4 % of all light jets is found. This is done by selecting only jets with fewer than three clusters, retaining 0.270 % of all uds-jets in the sample.

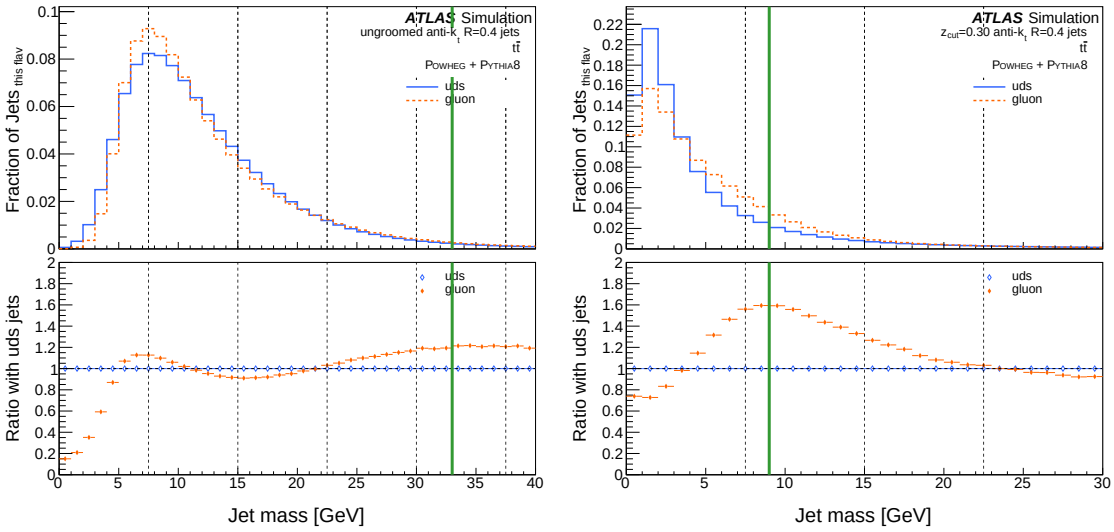
Similar results occur when cutting on jet mass to select uds-jets. The maximum purity that achieved is 89.1 % by selecting ungroomed jets with jet mass less than 1 GeV. This purity can't be improved upon through grooming, however this only keeps 0.0605 % of original uds-jets.

The opposite is true, however, for gluon-jets, where grooming steadily increases the maximum possible purity available. We also find that the remaining sample size, following the selection cut, increases. Firstly, the effects of soft drop grooming on choosing selection cuts for NCLUS were investigated. At ungroomed level the best result occurs when removing all jets with



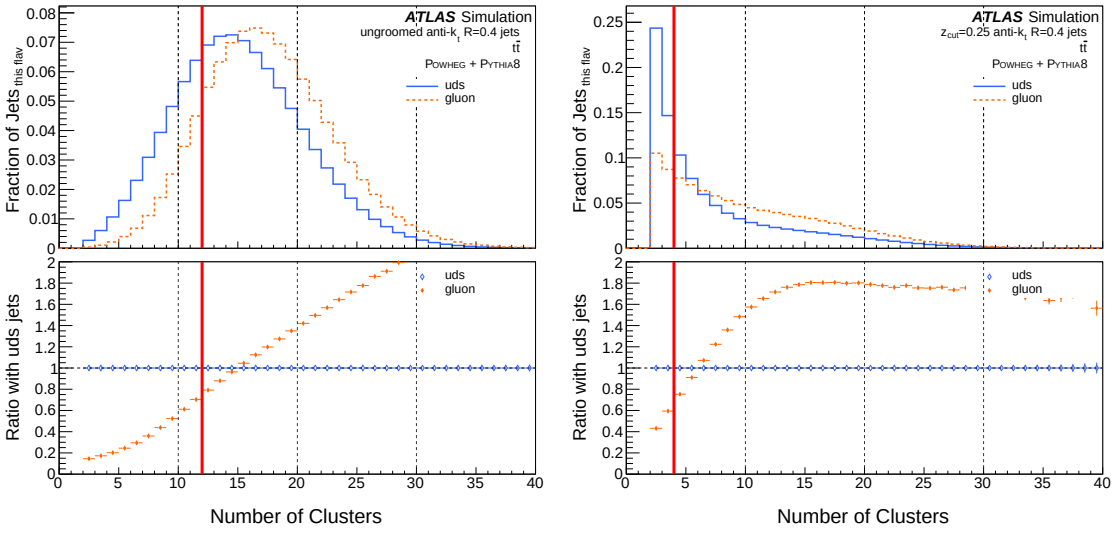
(a) ungroomed (b) $Z_{cut} = 0.3$

Figure 6.32: NCLUS for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.3$, (b), light flavour-tagged jets, normalised to the number of uds-jets with a ratio to uds-jets. The green line indicates the selection cut to maximise gluon-jets.



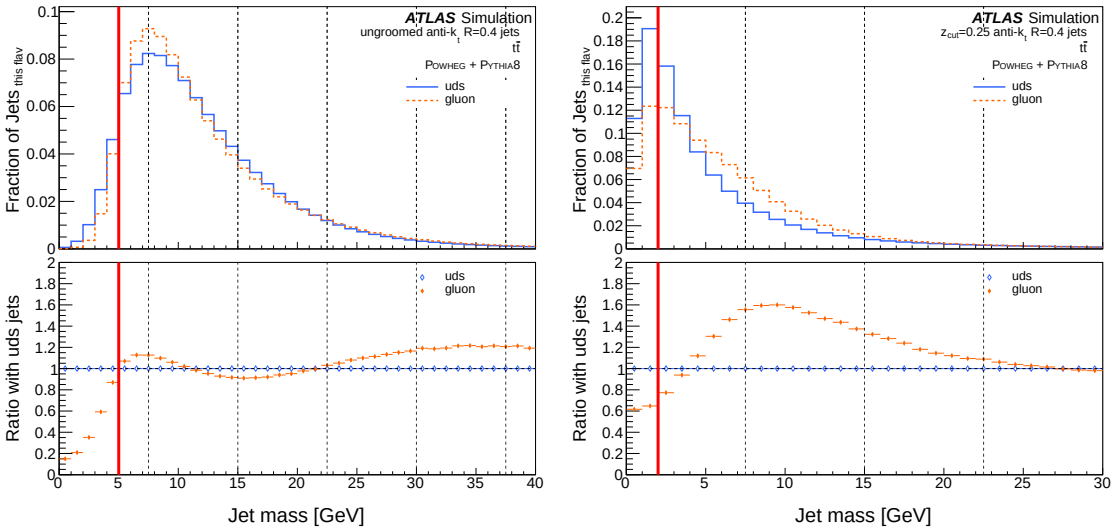
(a) ungroomed (b) $Z_{cut} = 0.3$

Figure 6.33: Jet mass for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.3$, (b), light flavour-tagged jets, normalised to the number of uds-jets with a ratio to uds-jets. The green line indicates the selection cut to maximise gluon-jets.



(a) ungroomed (b) $Z_{cut} = 0.25$

Figure 6.34: NCLUS for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.25$, (b), light flavour-tagged jets, normalised to the number of uds-jets with a ratio to uds-jets. The red line indicates the selection cut with an 80 % WP to maximise gluon-jets.



(a) ungroomed (b) $Z_{cut} = 0.25$

Figure 6.35: Jet mass for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.25$, (b), light flavour-tagged jets, normalised to the number of uds-jets with a ratio to uds-jets. The red line indicates the selection cut with an 80 % WP to maximise gluon-jets.

fewer than 38 clusters. This gives a sample purity of 66.4 %, but this only retains 0.0559 % of all original gluon-jets; however, at the maximum soft drop grooming strength of $Z_{cut} = 0.30$ a purity of 69.7 % is reached by removing all jets with fewer than thirteen clusters. This selection keeps 25.8 % of gluon-jets in the sample. These cuts can be seen in Figure 6.32

Moving onto selection cuts that satisfy the 80 % WP similar results arefound for the purity attainable from the light-jets. Firstly, uds-jets the purity of the selected samples cannot be improved through the use of soft drop grooming. For NCLUSthe best attainable purity, whilst retaining at least 80 % of the sample size, is accomplished by selecting only ungroomed jets with fewer than 20 clusters. This gives a purity of 59.0 % and preserves 81.2 % of all uds-jets. For selection cuts applied to jet mass of uds-jets, the best attainable purity comes from the ungroomed jets. We accomplish 55.3 % purity by selecting only jets with jet mass less than 21 GeV, keeping 91.0 % of the uds-jets.

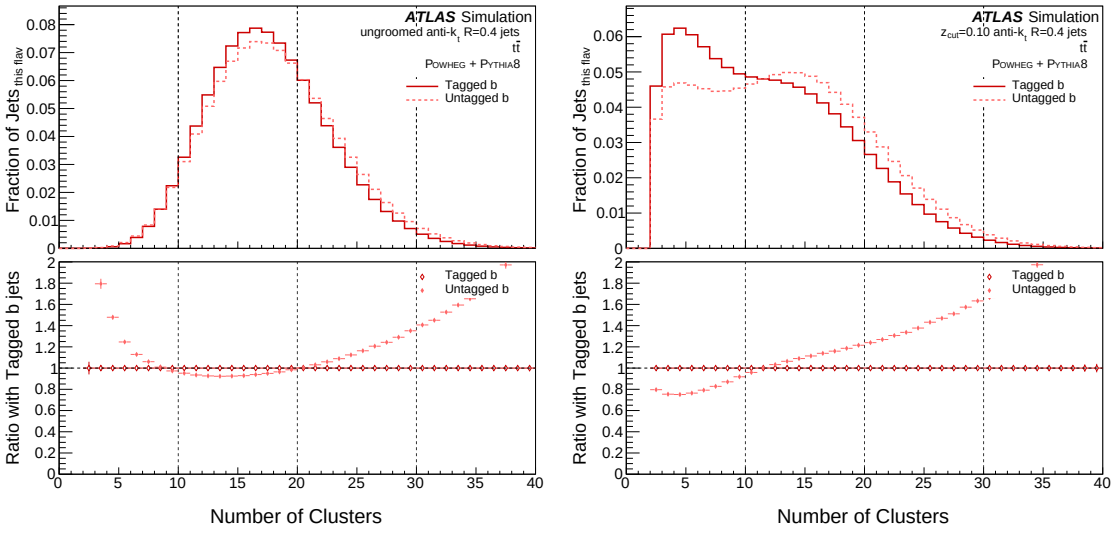
For gluon-jets the story is once again that of increasing purity by using soft drop grooming before applying selection cuts. By only selecting ungroomed gluon-jets with at least twelve clusters a purity of 49.8 % can be reached, whilst still maintaining 85.2 % of the sample jets. Following application of soft drop grooming with strength $Z_{cut} = 0.25$, and selecting only jets with at least four clusters, 61.4 % sample purity can be obtained, whilst still keeping 80.8 % of the gluon-jets. This is shown in Figure 6.33.

Selecting jet mass cuts on gluon-jets to maximise purity, whilst still retaining 80 %, also improves following soft drop grooming. Applying a cut on ungroomed jets that removes any with jet massless than 5 GeV gives a purity of 45.6 %. This keeps 94.1 % of gluon-jets. This is improved upon by applying soft drop grooming the jets before implementing the cut. Using jets that have first had soft drop grooming applied, with a strength of $Z_{cut} = 0.25$, a purity of 52.3 % can be achieved by removing all jets with a jet mass less than 2 GeV. This allows me to keep 80.7 % of the gluon-jets. This is shown in Figure 6.34

Over all, applying soft drop grooming is successful for improving possible sample purity for gluons, whether restricting this to reaching an 80 % WP of gluon-jets or not. The different behaviour displayed by uds- and gluon-jets following the application of soft drop grooming – particularly the more dramatic changes in uds-jets from applying soft drop grooming: the shift in distribution shape of NCLUS, the bigger drop in peak jet mass and the larger reduction in sample jets - means that grooming the jets improves the ability for separation of light-jets from one another. This is shown in Figure 6.35

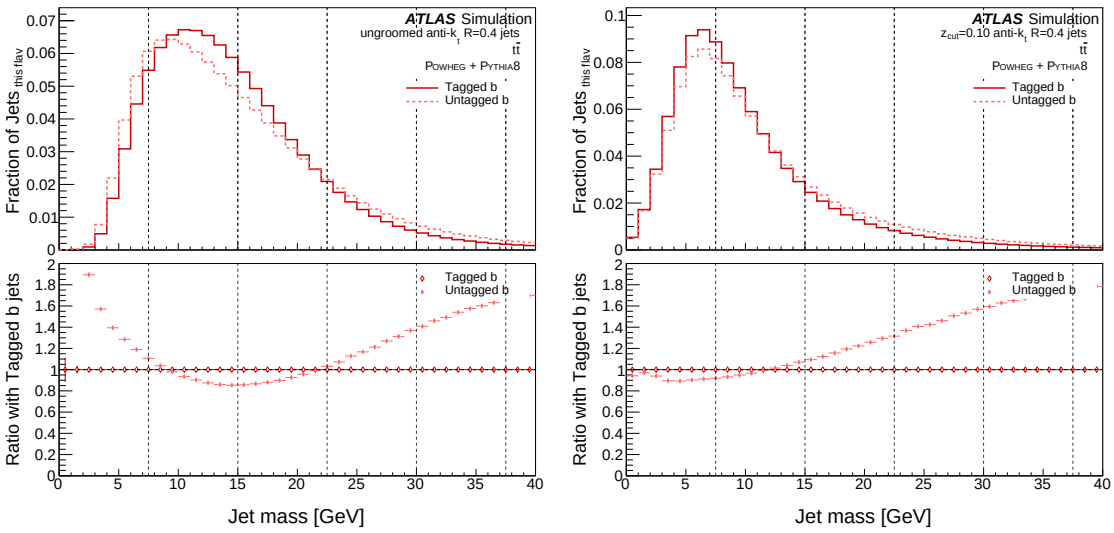
6.9 Soft Drop Grooming and b-Tagging

We can deepen the understanding of the effects of soft drop grooming on b-jets by investigating any potential differences between those b-jets identified by traditional tagging methods (See Section 6.4.3), and those which remain “untagged”. Understanding common or different behaviours of these b-jets, and how untagged b-jets compare to non-b-jets, may allow me to improve identification methods.



(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.36: NCLUS for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), *b*-tagged flavour-tagged jets with a ratio to *b*-tagged jets.



(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.37: Jet mass of Ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), *b*-tagged flavour-tagged jets with a ratio to *b*-tagged jets.

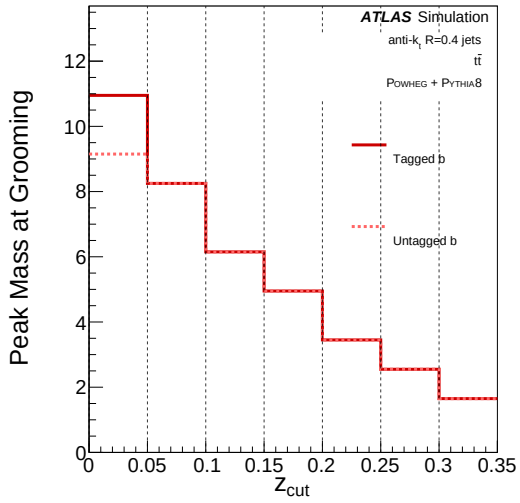


Figure 6.38: The peaks of the jet mass distributions for tagged- and untagged-b-jets at different strengths of soft drop grooming.

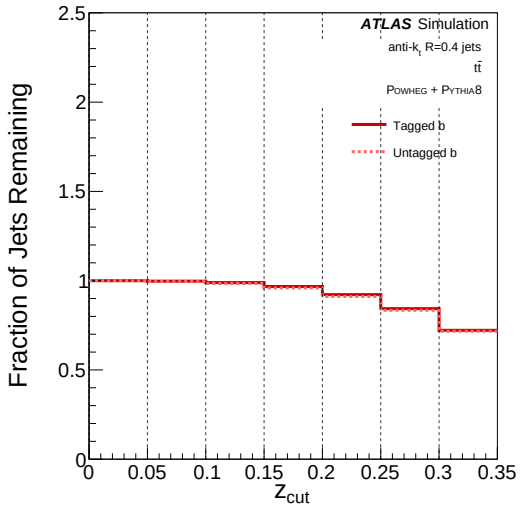


Figure 6.39: The fraction of tagged- and untagged-b-jets remaining following different strengths of soft drop grooming.

Firstly I can look make comparisons between tagged b-jets and untagged b-jets. Figure 6.36 shows NCLUS for tagged b-jets and untagged b-jets, both ungroomed and with soft drop grooming strength $Z_{cut} = 0.$ applied. We display a ratio plot beneath, to compare tagged and untagged b-jets within each bin in closer detail. When comparing the ungroomed jets I see a strong overlap with only slight differences. Both tagged b-jets and untagged b-jets peak at sixteen clusters, but the distribution for untagged b-jets is very slightly broader, and slightly shifted to higher values of NCLUS. Afer using soft drop grooming I start to see tagged b-jets and untagged b-jets behaving differently. Both are starting to develop a second, earlier peak, at four clusters, close to the minimum value of two, however this is certainly more prominent for the tagged b-jets, where this new peak is the highest, unlike for untagged b-jets. NCLUS for all grooming strengths are shown in Figure D.3.

Next, I can compare the results for jet mass, in Figure 6.37. When the jets are ungroomed I see a similar distribution between tagged b-jets and untagged b-jets, peaking at 10 GeV and 9 GeV respectively. The shape of the distribution is a little broader for untagged b-jets, and shifted to slightly lower masses. Following grooming, the two distributions are even closer than before. Both now peak at 6 GeV, but untagged b-jets still have a slightly broader distribution and are slightly shifted to heavier jet mass. jet mass for all grooming strengths are shown in Figure D.4. We also see this present in Figure 6.38, with the peak mass of tagged b-jets starting off about 20 % higher in ungroomed jets, and quickly dropping down to matchuntagged b-jets following grooming.

The combination of these features before soft drop grooming suggests initially that tagged b-jets, with slightly fewer clusters and slightly higher values of jet mass, may have a more centralised structure, with a harder core than untagged b-jets. The effects of grooming are more dramatic on the value of NCLUS than on jet mass for both tagged b-jets and untagged b-jets, and suggests, as expected, the clusters I are grooming away are softer. This effects of grooming, and in this behaviour in particular, is more apparent in tagged b-jets, however. The more apparent effect of soft drop grooming on tagged b-jets indicates more initial definition between hard and soft clusters, as the higher number of clusters, and smaller value of jet mass lost, show these clusters are softer. Following the first soft drop grooming strength, $Z_{cut} = 0.05$, the mass of both tagged b-jets and untagged b-jets remains very close, indicating these many, soft clusters in tagged b-jetss were groomed away easily, and the pattern of tagged b-jets with fewer clusters with more jet mass remains throughout.

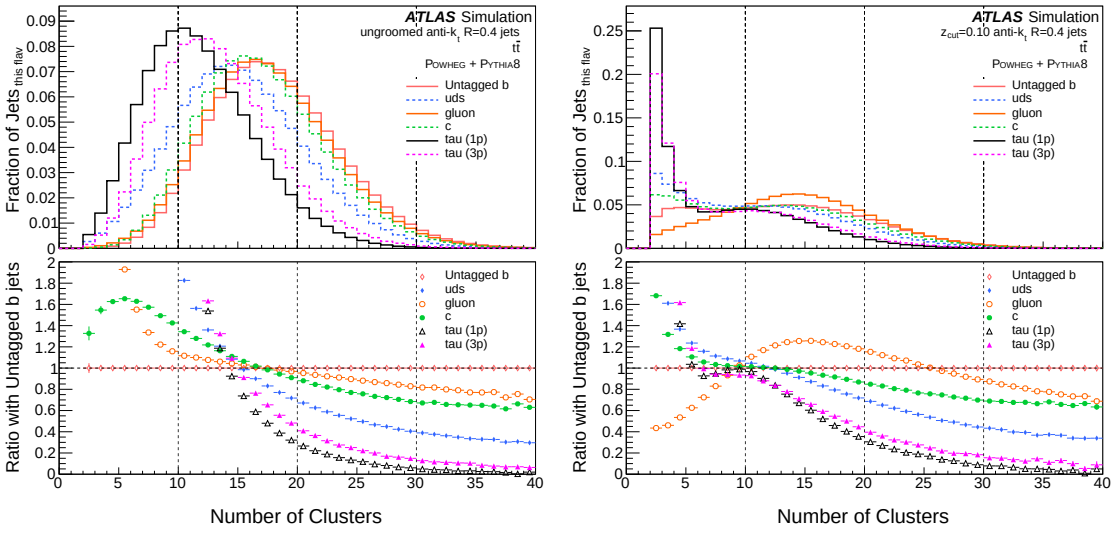
When comparing the portion of the sample lost to soft drop grooming I see very little difference between tagged b-jets and untagged b-jets, with very slight fewer untagged b-jets lost. This is shown in Figurer 6.39. This could relate to tagged b-jets having fewer clusters than untagged b-jets, and thus more likely to be groomed below the minimum NCLUS value of two clusters.

Selection Cuts on untagged bjets

Can any of the common, distinct features of b-jets be used to improve identification of untagged b-jets against a background of non-b-jets? Comparing the observables of untagged b-jets to all non-b-jet flavours may answer this question. Firstly, I present the results for NCLUS for these sets of jets in Figure 6.40.

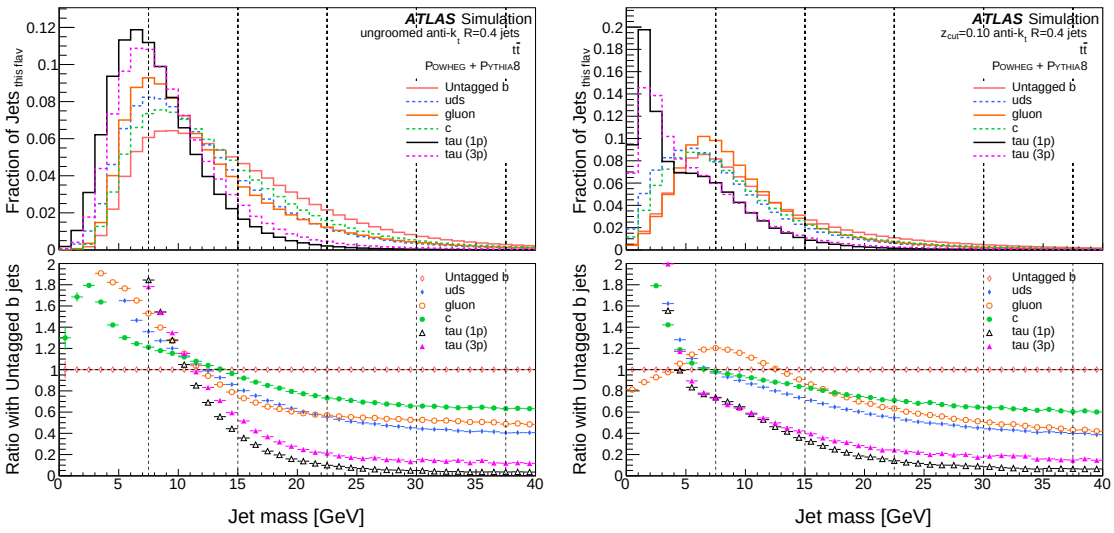
At the ungroomed level the distribution seen is much like that for all b-jets, of all the flavours shown it has the highest value for NCLUS and a slightly broader distribution. At groomed level there is only a slight difference, with untagged b-jets affected by soft drop grooming slightly less than the group containing all b-jets, but still following a similar trend: other than gluon-jets, it still has the highest and broadest distribution for NCLUS.

Following this I present jet mass for the untagged b-jets. These are shown in Figure 6.41. Again, at ungroomed level the result is much the same as the group containing all b-jets: the



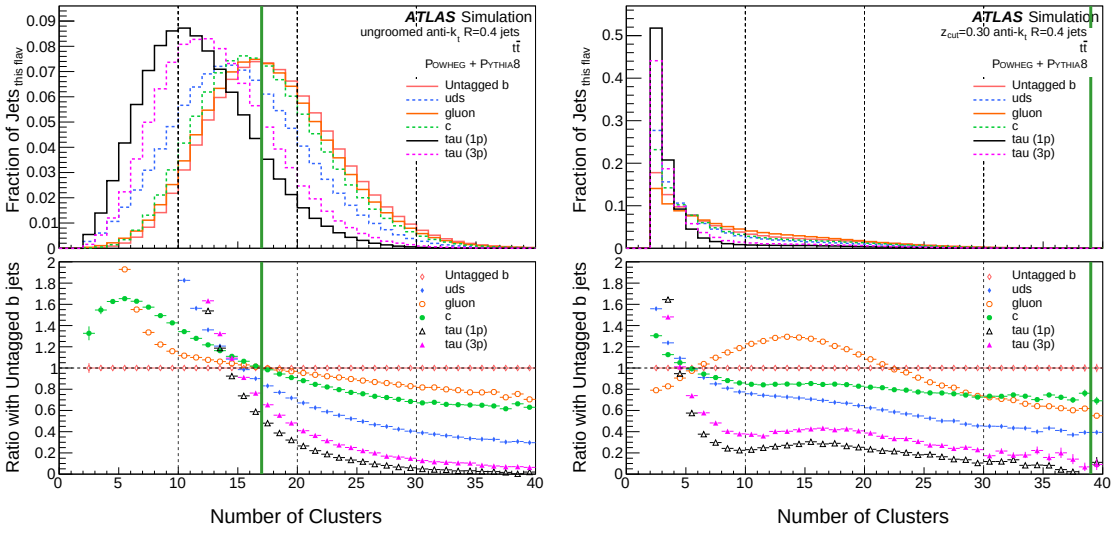
(a) ungroomed (b) $Z_{\text{cut}} = 0.1$

Figure 6.40: NCLUS for ungroomed, (a), and soft drop groomed with strength $Z_{\text{cut}} = 0.10$, (b), untagged-b-jets and non-b-flavour-untagged jets with a ratio to untagged-b jets.



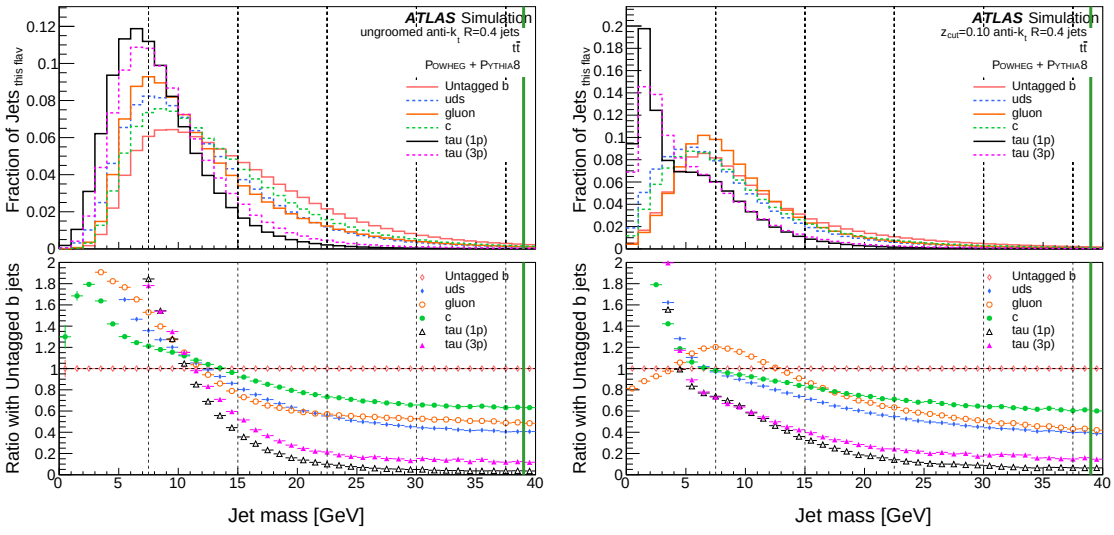
(a) ungroomed (b) $Z_{\text{cut}} = 0.1$

Figure 6.41: Jet mass for ungroomed, (a), and soft drop groomed with strength $Z_{\text{cut}} = 0.10$, (b), untagged-b-jets and non-b-flavour-untagged jets with a ratio to untagged-b jets.



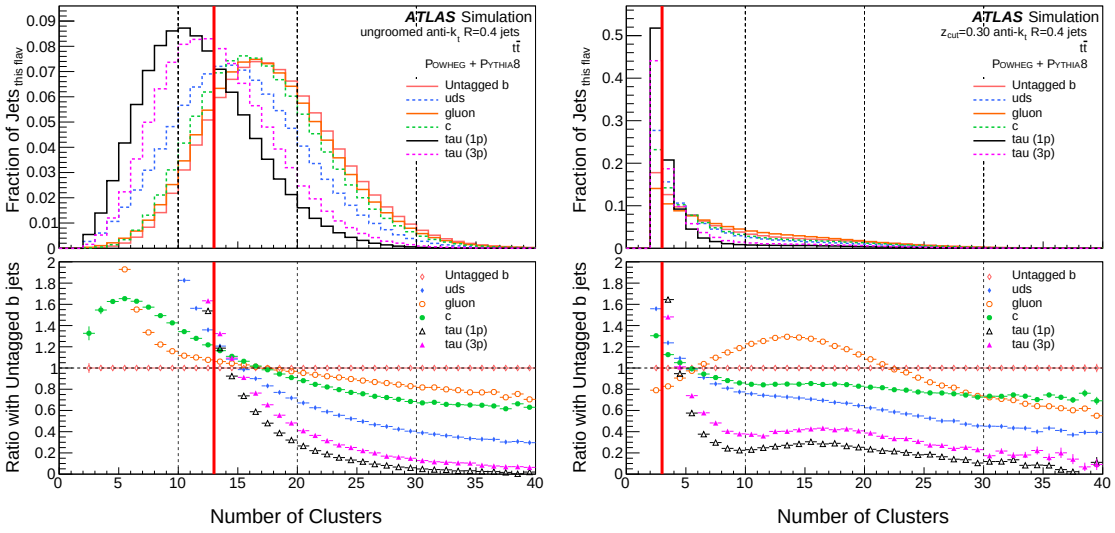
(a) ungroomed (b) $Z_{cut} = 0.3$

Figure 6.42: NCLUS for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.3$, (b), untagged-b-jets and non-b-flavour-untagged jets with a ratio to untagged-b jets. The green line indicates the selection cut.



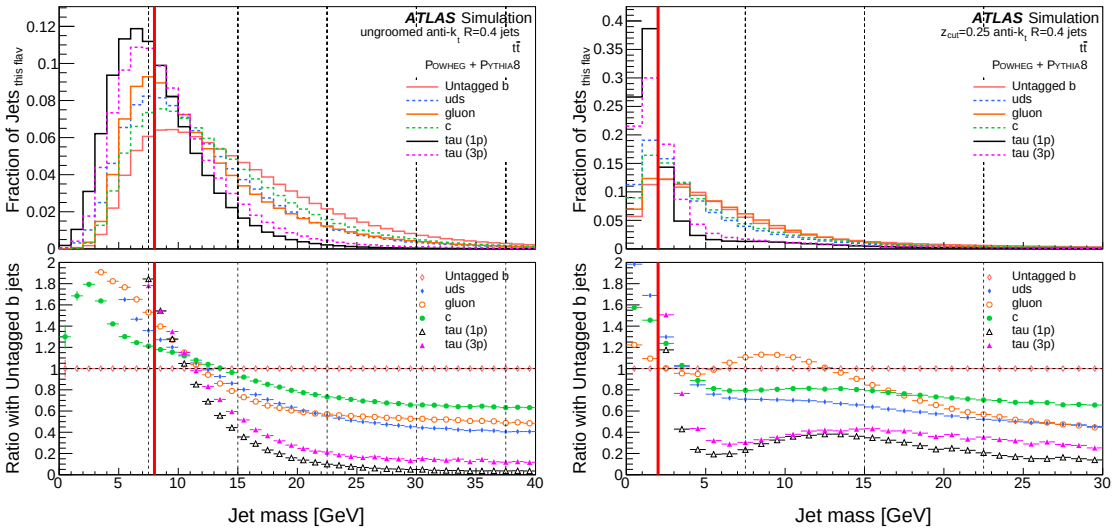
(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.43: Jet mass for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), untagged-b-jets and non-b-flavour-untagged jets with a ratio to untagged-b jets. The green line indicates the selection cut.



(a) ungroomed (b) $Z_{cut} = 0.3$

Figure 6.44: NCLUS for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.3$, (b), untagged-b-jets and non-b-flavour-untagged jets with a ratio to untagged-b jets. The red line indicates the selection cut with an 80 % WP.



(a) ungroomed (b) $Z_{cut} = 0.25$

Figure 6.45: Jet mass for ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.25$, (b), untagged-b-jets and non-b-flavour-untagged jets with a ratio to untagged-b jets. The red line indicates the selection cut with an 80 % WP.

distribution for jet mass is the highest and broadest of all flavours. The effects on the jet mass results following the application of soft drop grooming is even more similar to all b-jets than it was for NCLUS: other than gluon-jets the distribution remains the highest and broadest. What is seen is that these untagged b-jets, like all b-jets, start off with a high value of NCLUS and jet mass, and are moderately effected by soft drop grooming – more so than gluon-jets but less so than tau-jets.

Following this result, I investigate the ability to produce a single-variable cut on these samples to improve signal purity. Starting with possible cuts on the value of NCLUS, I find that it is possible to improve the maximum purity of the sample following application of soft drop grooming. The maximum possible purity attainable at ungroomed level is 32.4% by only selecting those jets with at least 39 clusters. This retains 0.00801% of ungroomed untagged b-jets. This is improved by applying a soft drop grooming of $Z_{cut} = 0.30$ an selecting only jets with at least 39 clusters, giving 36.1% purity. This retains 0.00706% of groomed untagged b-jets, 0.00506% of the original sample. These cuts are displayed in Figure 6.42.

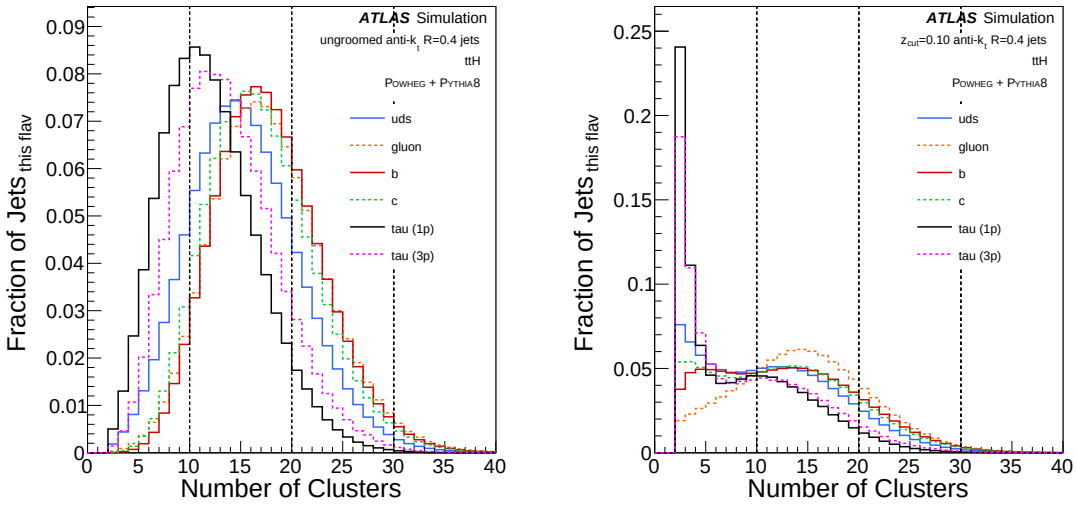
A similar result is true when investigating jet mass. The maximum possible purity attainable at ungroomed level is 33.5%, by selecting only those jets with a jet mass of at least 39 GeV. This retains 0.0803% of untagged b-jets. This can be improved by applying soft drop grooming with a strength of $Z_{cut} = 0.10$ and selecting only jets with a jet mass of at least 39 GeV, giving a purity of 37.7%. This retains 0.0757% of groomed untagged b-jets, 0.0748% of the original sample These cuts are displayed in Figure 6.43.

Next, I explore whether imposing a single variable cut on groomed jets can improve purity whilst retaining an efficiency of 80%. I start with results for NCLUS, where an improvement is indeed possible. At ungroomed level it is possible to obtain a purity of 21.4% whilst retaining 82.5% of untagged b-jets. This is done by selecting only those jets with at least 13 clusters. After applying soft drop grooming of strength $Z_{cut} = 0.30$, however, a purity of 23.0% can be achieved, whilst still keeping 82.2% of groomed untagged b-jets. This is done by selecting jets with at least three clusters, which retains 58.9The same is true once again for jet mass. The maximum possible purity achievable for ungroomed untagged b-jets is 21.2%, by keeping only jets with a jet mass of at least 8 GeV, maintaining 81.5% of untagged b-jets. Following an application of soft drop grooming with strength $Z_{cut} = 0.25$, a purity of 23.3% is reachable, by retaining only jets with a jet mass of at least 2 GeV. This keeps 83.0% of groomed untagged b-jets, 68.3% of the original sample. These cuts are displayed in Figure 6.45.

Overall it is seen that application of soft drop grooming can marginally improve the purity of untagged b-jets within the sample, even with the 80% efficient WP imposed.

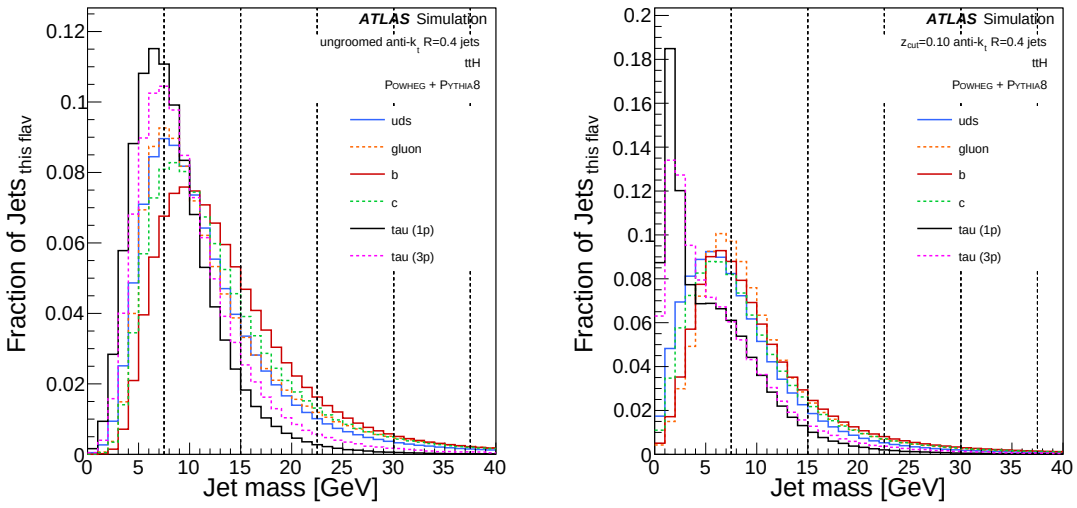
6.10 Soft Drop Grooming with $t\bar{t}H$ and $t\bar{t}$

So far in this chapter, all of the results have been JSS observables of flavour-tagged small jets from $t\bar{t}$ events. Do the changes I see here in JSS observables following the application of soft drop grooming reflect what I also see in flavour-tagged small jets from $t\bar{t}H$ events? If so, does



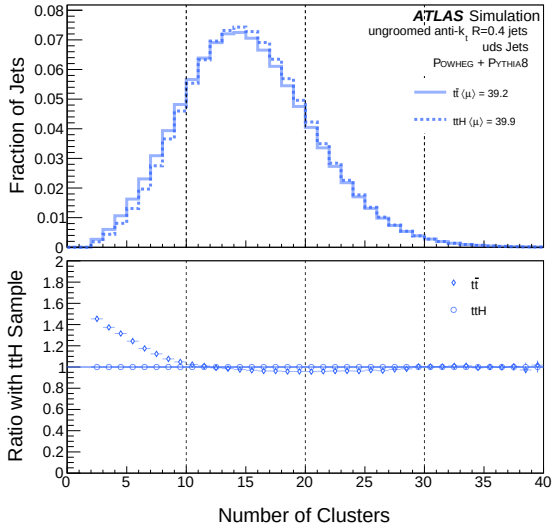
(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.46: NCLUS for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), flavour-tagged jets from $t\bar{t}H$ events.

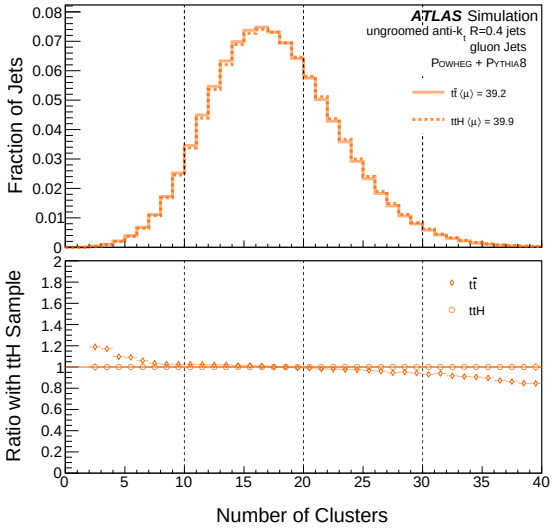


(a) ungroomed (b) $Z_{cut} = 0.1$

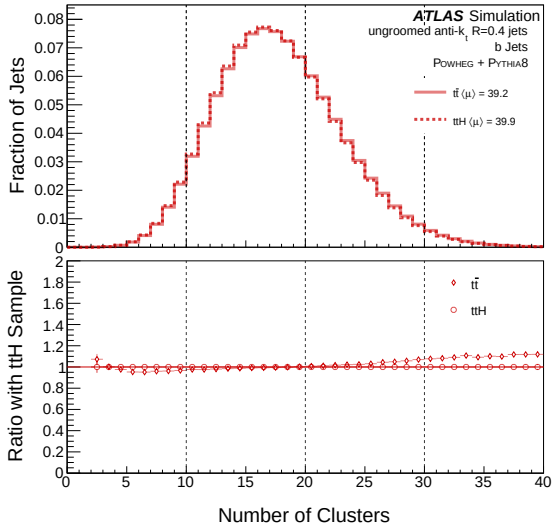
Figure 6.47: Jet mass for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), flavour-tagged jets from $t\bar{t}H$ events.



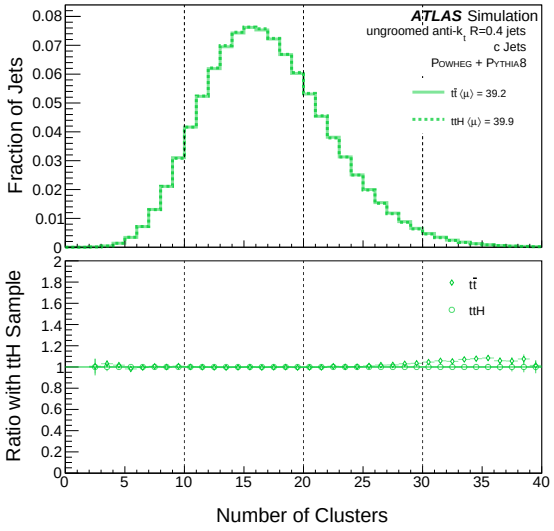
(a) uds



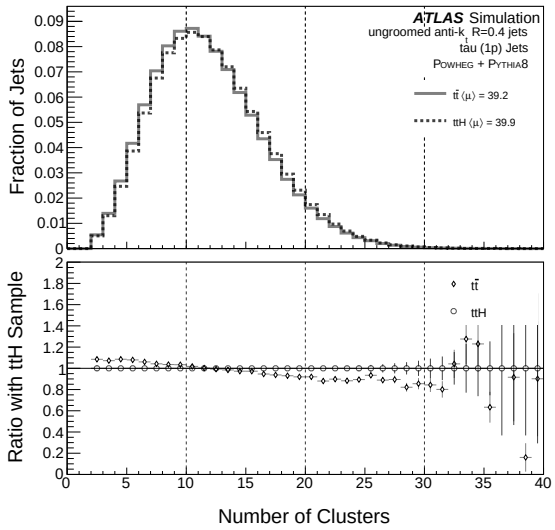
(b) gluon



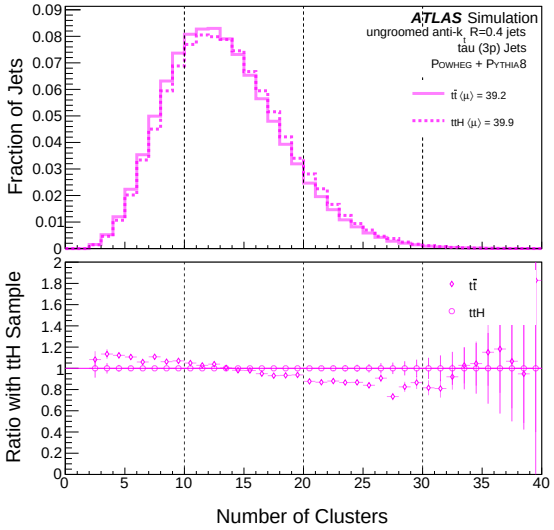
(c) b



(d) c

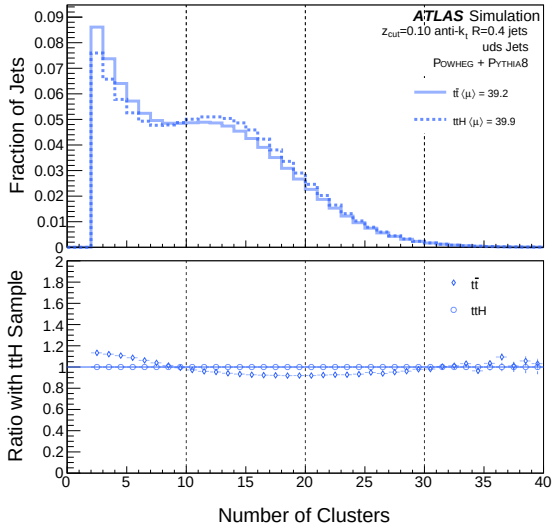


(e) 1-pronged tau

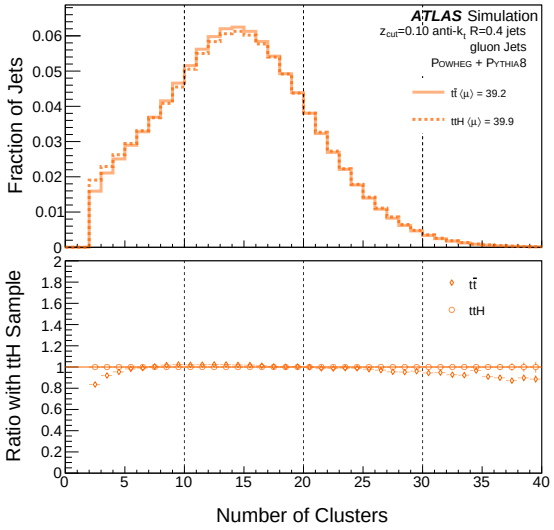


(f) 3-pronged tau

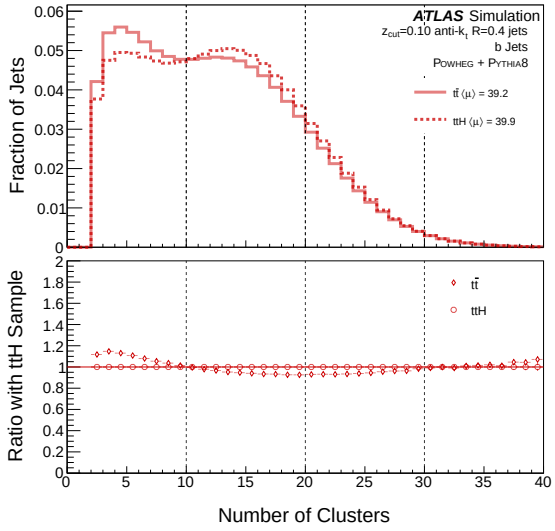
Figure 6.48: NCLUS for ungroomed $t\bar{t}$ and $t\bar{t}H$ jets of different flavours.



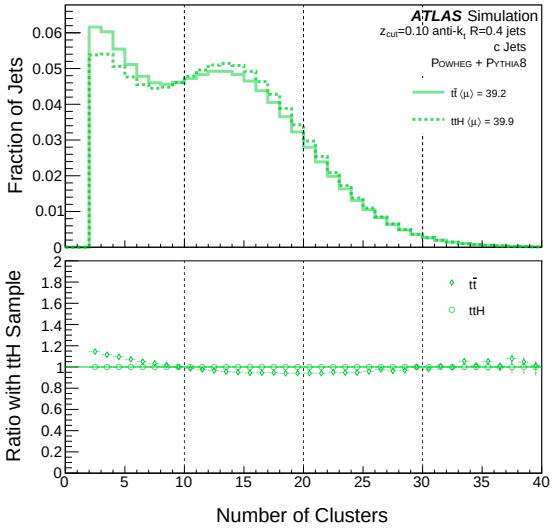
(a) uds



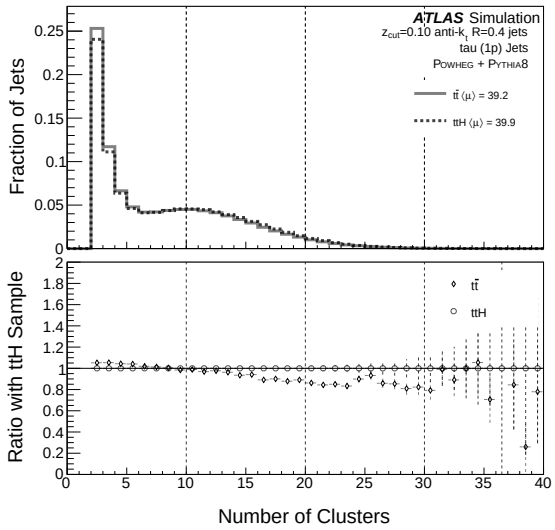
(b) gluon



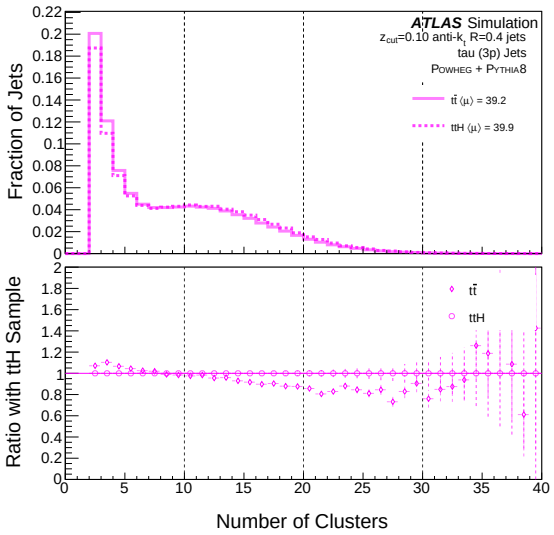
(c) b



(d) c



(e) 1-pronged tau



(f) 3-pronged tau

Figure 6.49: NCLUS for $t\bar{t}$ and $t\bar{t}H$ jets of different flavours, soft drop grooming with $Z_{cut} = 0.10$.

that lead to potential improvements in flavour-tagging jets in $t\bar{t}H$ events? If not, does that lead me to a potential method for improving separation between $t\bar{t}$ and $t\bar{t}H$ events?

We begin by investigating the NCLUS within jets in $t\bar{t}H$ events, with distributions shown in Figure 6.46 and Flavour-by-Flavour comparisons in Figure 6.48. We generally see similar distributions for jets within $t\bar{t}H$ to those from $t\bar{t}$ events. At ungroomed level, shown in Figure 6.48 there is very little distinction between jets from either of these events, however some differences do appear following the application of soft drop grooming, with the default strength of $Z_{cut} = 0.10$, shown in Figure 6.49. Here some differences in shape between $t\bar{t}$ and $t\bar{t}H$ are present, most noticeably for b-jets, but also in c-jets and uds-jets. In these jet flavours I see that jets from $t\bar{t}$ samples appear to be affected by the process of soft drop grooming, with the development of an early peak at a low value of NCLUS that is more prominent for jets in $t\bar{t}$ events. This indicates that at this strength more clusters are groomed away within jets in $t\bar{t}$ events than jets of the same flavour in $t\bar{t}H$.

Moving onto jet mass I notice again that generally the distributions are very close, with distributions shown in Figure 6.47 and Flavour-by-Flavour comparisons in Figure E.8. For most flavours of ungroomed jets there is a slight difference in the distributions of jets from $t\bar{t}$ and $t\bar{t}H$ samples, where I observe a higher peak for $t\bar{t}H$ samples, and for $t\bar{t}$ a there is a slightly broader distribution, skewed to slight higher values for jet mass. This is particularly prominent for b-jets, least for tau-jets, and seemingly non-existent for gluon-jets. Following soft drop grooming of default strength $Z_{cut} = 0.10$ I see these differences are heavily reduced and the distributions are now much closer.

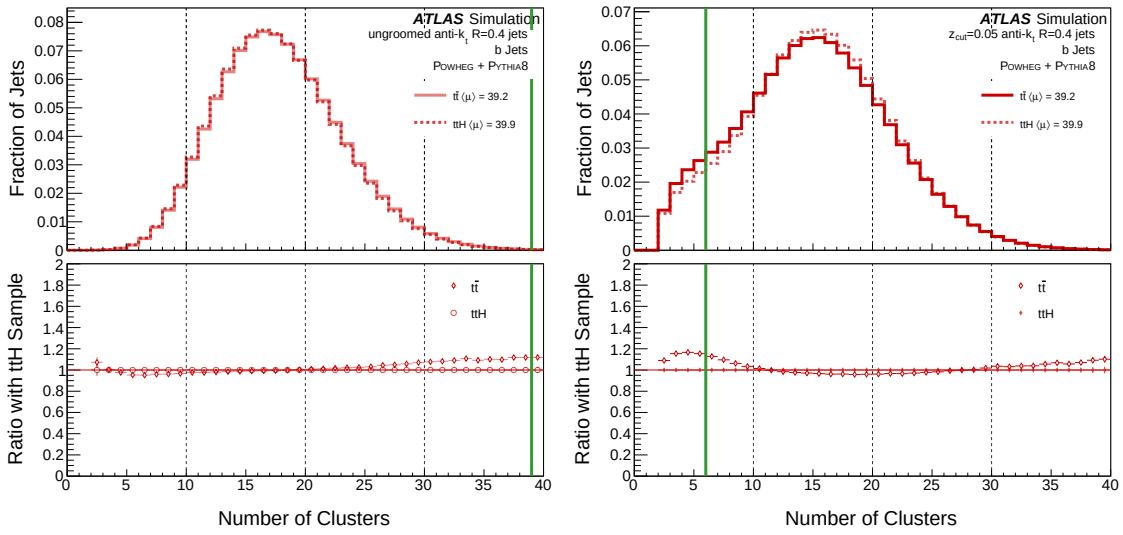
Selection Cuts for $t\bar{t}H$ and $t\bar{t}$ samples

Can I use soft drop grooming to improve the capability of selection cuts to increase the purity of the sample? Here I compare how these selection cuts can increase separation between b-jets, and b-tagged b-jets, from $t\bar{t}$ and $t\bar{t}H$ events. Can imposing selections cuts on groomed jets do a better job at separating jets from $t\bar{t}$ and $t\bar{t}H$ than selection cuts imposed on ungroomed jets? The number of jets are normalised to the number of $t\bar{t}$ and $t\bar{t}H$ events.

Firstly I look at b-jets from both sets of events. The highest possible purity of $t\bar{t}$ I can acquire from selection cuts on ungroomed jets is 59.5 % by removing jets with NCLUS fewer than 39, however from b-jets that have been had a soft drop grooming strength of $Z_{cut} = 0.05$ I can achieve $t\bar{t}$ purity of 60.1 % by only keeping jets with fewer than 6 clusters. This is shown in Figure 6.50.

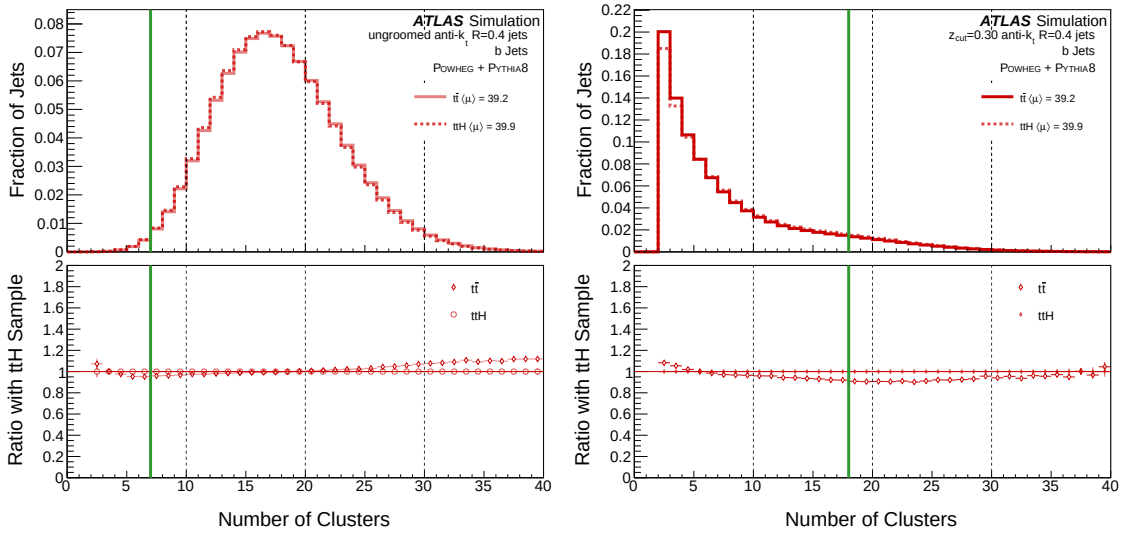
For ungroomed jets I can attain 44.3 % purity for $t\bar{t}H$ b-jets by only keeping jets with fewer than 7 clusters. Using jets with a soft drop grooming of strength $Z_{cut} = 0.30$ I can get purity of 45.4 % by removing all jets with fewer than 18 clusters. We see that grooming allows me to improve purity for both sample sets in this way. This is shown in Figure 6.51.

If I want to impose selection cuts that only retain 80 % of $t\bar{t}H$ b-jets, I still find I are able to increase $t\bar{t}$ b-jet purity by using jets with soft drop grooming applied. Using ungroomed jets I



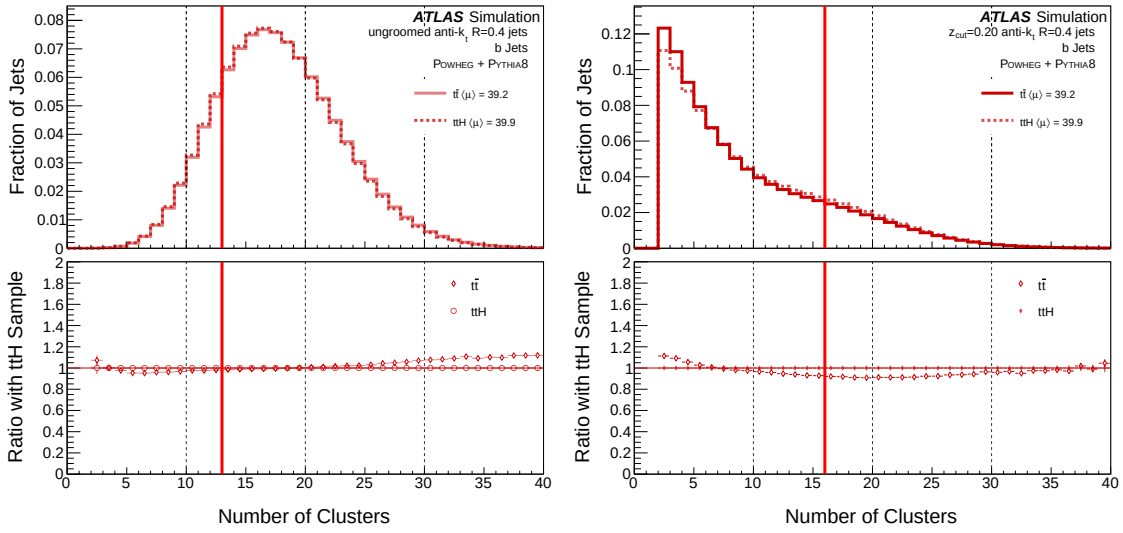
(a) ungroomed (b) $Z_{cut} = 0.05$

Figure 6.50: NCLUS for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.05$, (b), flavour-tagged jets from $t\bar{t}$ and $t\bar{t}H$ events. The green line indicates the selection cut to maximise $t\bar{t}$ events.



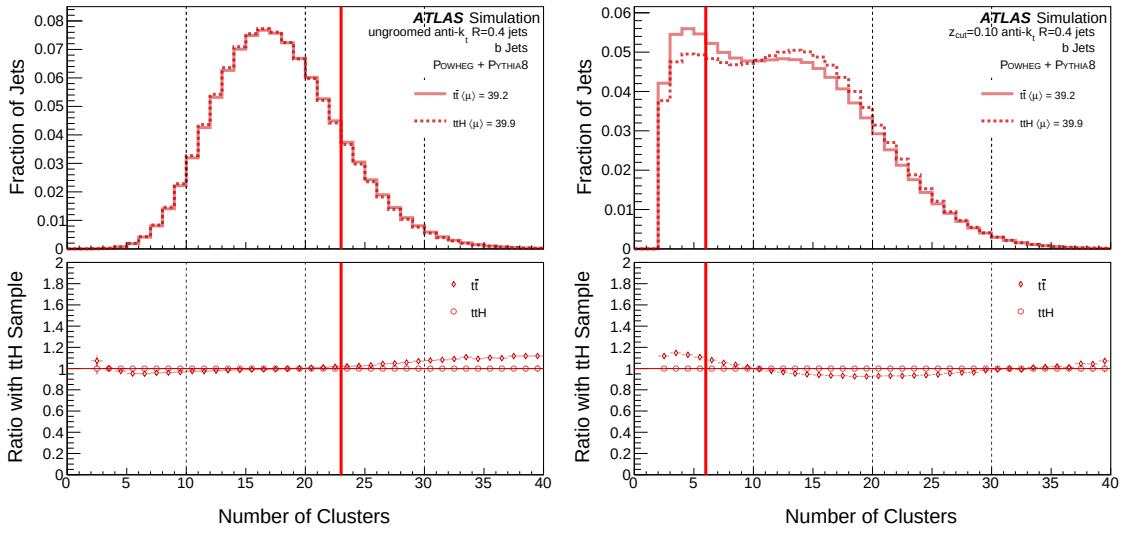
(a) ungroomed (b) $Z_{cut} = 0.3$

Figure 6.51: NCLUS for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.3$, (b), flavour-tagged jets from $t\bar{t}$ and $t\bar{t}H$ events. The green line indicates the selection cut to maximise $t\bar{t}H$ events.



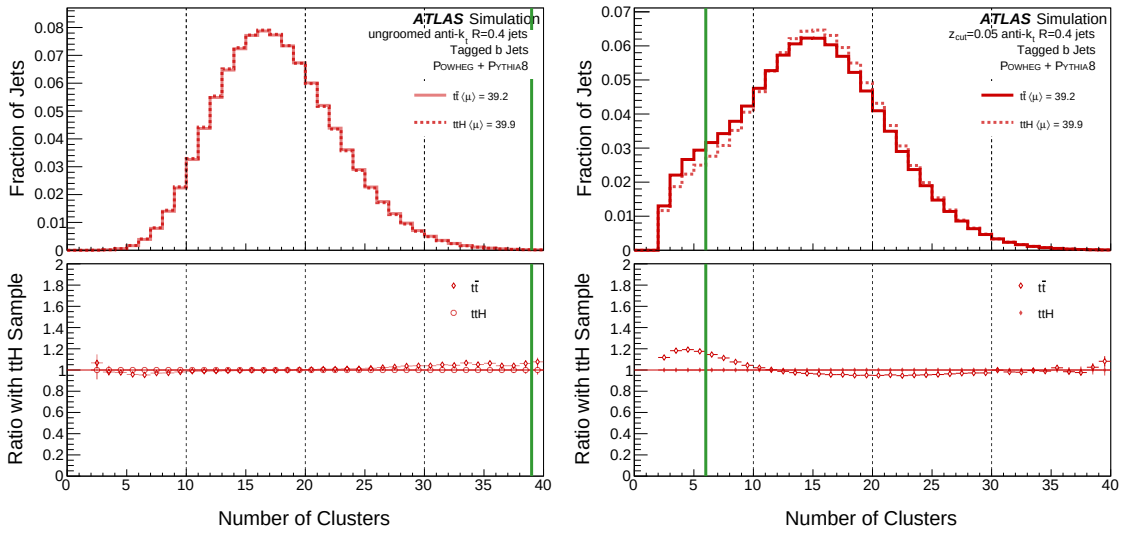
(a) ungroomed (b) $Z_{cut} = 0.2$

Figure 6.52: NCLUS for the ungroomed, (a), and soft drop groomed with $Z_{cut} = 0.2$, (b), flavour-tagged jets from $t\bar{t}$ and $t\bar{t}H$ events. The red line indicates the selection cut with an 80 % WP to maximise $t\bar{t}$ events.



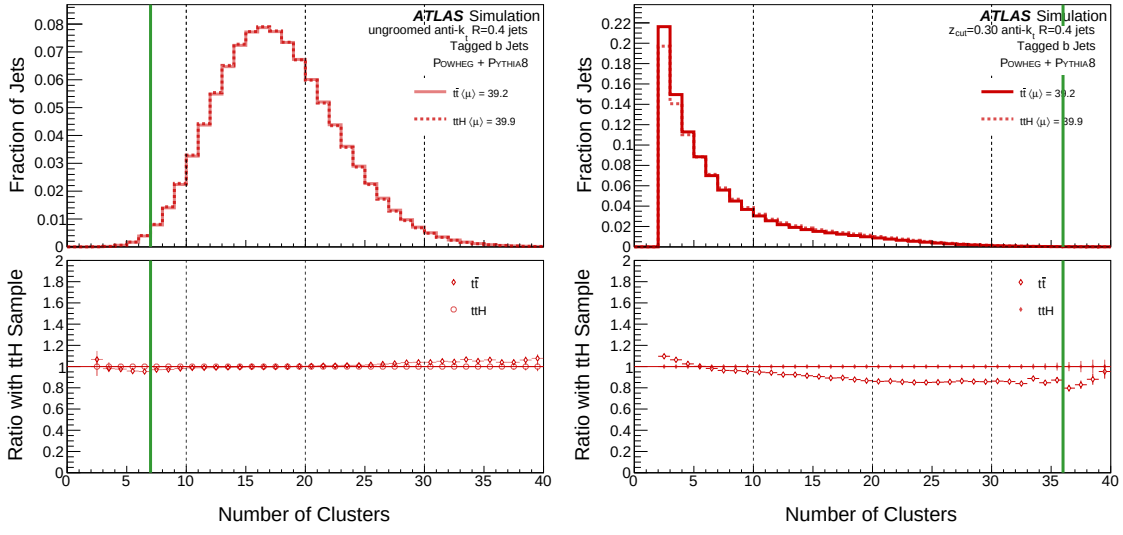
(a) ungroomed (b) $Z_{cut} = 0.1$

Figure 6.53: NCLUS for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.10$, (b), flavour-tagged jets from $t\bar{t}$ and $t\bar{t}H$ events. The red line indicates the selection cut with an 80 % WP to maximise $t\bar{t}H$ events.



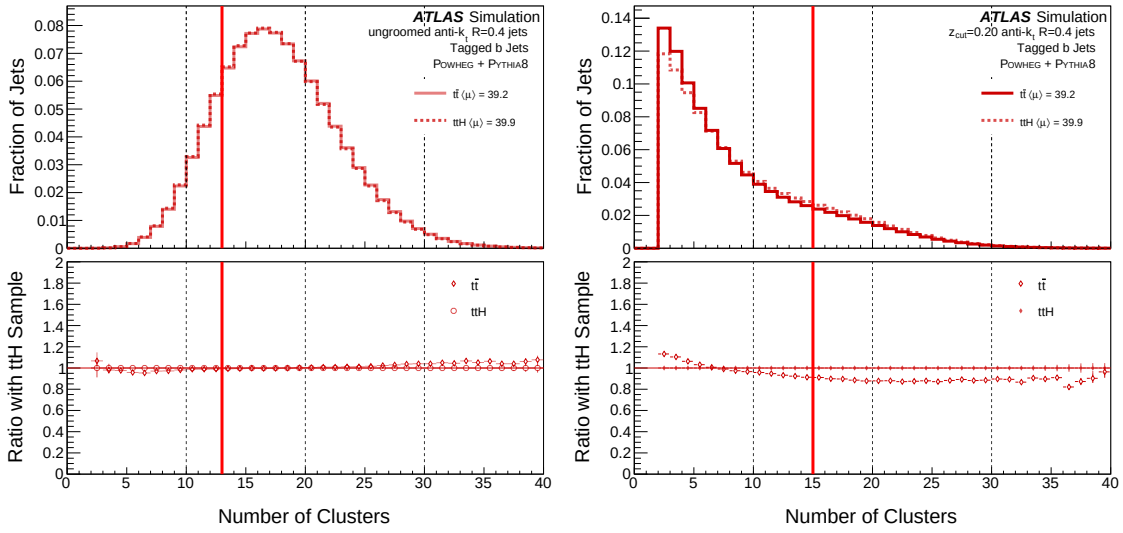
(a) ungroomed (b) $Z_{cut} = 0.05$

Figure 6.54: Number of clusters for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.05$, (b), flavour-tagged and b-tagged jets from $t\bar{t}$ and $t\bar{t}H$ events. The green line indicates the selection cut to maximise $t\bar{t}$ events.



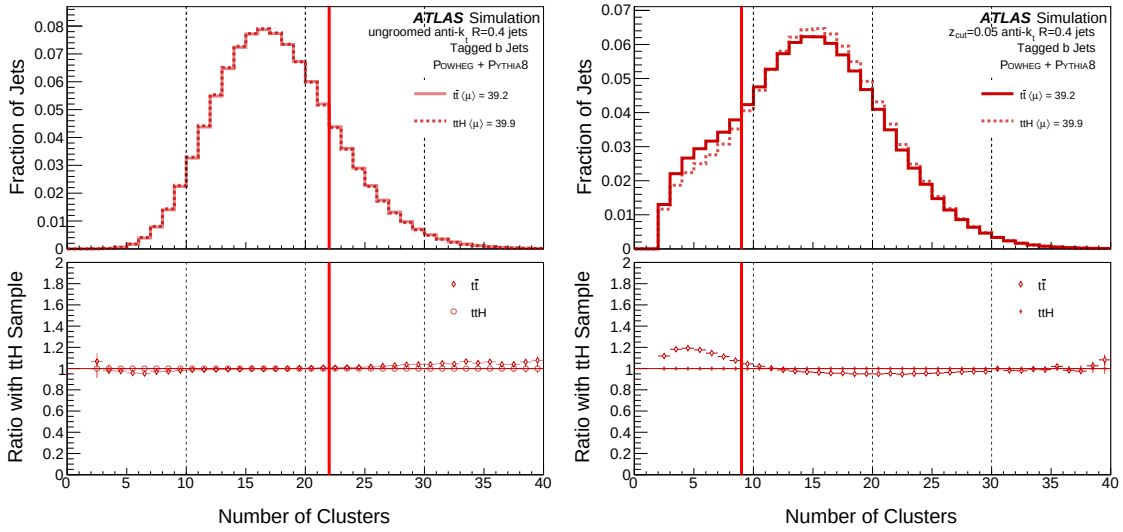
(a) ungroomed (b) $Z_{cut} = 0.3$

Figure 6.55: Number of clusters for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.3$, (b), flavour-tagged and b-tagged jets from $t\bar{t}$ and $t\bar{t}H$ events. The green line indicates the selection cut to maximise $t\bar{t}H$ events.



(a) ungroomed (b) $Z_{cut} = 0.2$

Figure 6.56: Number of clusters for the ungroomed, (a), and soft drop groomed with strength $Z_{cut} = 0.2$, (b), flavour-tagged and b-tagged jets from $t\bar{t}$ and $t\bar{t}H$ events. The red line indicates the selection cut with an 80 % WP to maximise $t\bar{t}$ events.



(a) ungroomed (b) $Z_{cut} = 0.05$

Figure 6.57: Number of clusters for the ungroomed, (a), and soft drop groomed with $Z_{cut} = 0.05$, (b), flavour-tagged and b-tagged jets from $t\bar{t}$ and $t\bar{t}H$ events. The red line indicates the selection cutto maximise $t\bar{t}H$ events.

can obtain a $t\bar{t}$ purity of 56.9 % by removing jets with fewer than 13 clusters. This keeps 81.6 % of groomed $t\bar{t}H$ b-jets, and 82.1 % of $t\bar{t}$ b-jets. Using jets with a soft drop grooming strength of $Z_{cut} = 0.20$ I can achieve a $t\bar{t}$ b-jet purity of 57.3 %. This is done by only keeping jets with fewer than 16 clusters. This retains 80.4 % of groomed $t\bar{t}H$ b-jets and 82.0 % of groomed $t\bar{t}$[] b-jets. This is shown in Figure 6.52.

If I want to keep 80 % of all soft drop grooming $t\bar{t}H$ b-jets than I can also improve purity. for ungroomed jets I can reach $t\bar{t}H$ b-jet purity of 43.4 % by only keeping jets with fewer than 23 clusters. Keeps 84.2 % of $t\bar{t}H$, and 83.6 % of $t\bar{t}$. With jets groomed with strength $Z_{cut} = 0.10$ I can accomplish a purity of 43.9 % by removing jets with fewer than 6 clusters. This keeps 81.6 % of groomed $t\bar{t}H$ b-jets and 79.3 % of groomed $t\bar{t}$ b-jets. Here I find that grooming allows me to improve purity for both sample sets, whilst still retaining an efficiency of 80 % for the $t\bar{t}H$ b-jets. This is shown in Figure 6.53.par

We can repeat this process using only b-jets that have been tagged through traditional identification methods, tagged b-jets. When comparing tagged b-jets from $t\bar{t}$ and $t\bar{t}H$ events I also see that applying selection cuts to jets that have had soft drop grooming applied can improve the purity of the sample more than applying selection cuts to ungroomed jets. The highest purity for $t\bar{t}$ tagged b-jets I can reach is 58.6 %, by only keeping jets with at least 39 clusters. We can improve on this by applying soft drop grooming strength of $Z_{cut} = 0.05$ and selecting jets with fewer than six clusters. – giving me a purity of $t\bar{t}$ tagged b-jets of 60.7 %. This is shown in Figure 6.54.

The maximum purity for $t\bar{t}H$ tagged b-jets I can attain is 44.3 %, by selecting only jets with fewer than 7 clusters. After applying soft drop grooming of strength $Z_{cut} = 0.30$, and selecting jets with at least 36 clusters, I can acquire a purity for $t\bar{t}H$ tagged b-jets of 47.5 %. This is shown in Figure 6.55.

If I once more set the requirement of keeping 80 % of the groomed $t\bar{t}H$ tagged b-jets following the selection cuts, I still find that applying soft drop grooming improves the effect of the selection cut in maximising purity. for ungroomed jets the maximum purity I can achieve for $t\bar{t}$ tagged b-jets, whilst retaining 80 % of $t\bar{t}H$ tagged b-jets, is 56.9 %. This is done by removing all jets with fewer than 13 clusters and retains 81.6 % of all $t\bar{t}H$ tagged b-jets and 81.8 % of $t\bar{t}$ tagged b-jets. Following the application of soft drop grooming strength $Z_{cut} = 0.20$ I are able to reach a $t\bar{t}$ tagged b-jets purity of 57.4 %. This is achieved through keeping only jets with fewer than 15 clusters. This selection cut keeps 80.5 % of $t\bar{t}H$ tagged b-jets and 82.7 % of $t\bar{t}$ tagged b-jets. This is shown in Figure 6.56.

We also see that application for soft drop grooming can improve sample purity for $t\bar{t}H$ tagged b-jets. The maximum attainable purity of $t\bar{t}H$ tagged b-jets using ungroomed jets is 43.3 %. This is by applying a selection cut that keeps only jets with fewer than 22 clusters, retaining 80.8 % of all $t\bar{t}H$ tagged b-jets and 80.5 % f $t\bar{t}$ tagged b-jets. Following soft drop grooming of strength $Z_{cut} = 0.05$ I can reach a purith for $t\bar{t}H$ tagged b-jets of 43.9 % by removing all jets with fewer than nine clusters. This keeps 82.9 % of $t\bar{t}H$ tagged b-jetsand 80.5 % of $t\bar{t}$ tagged b-jets. This is shown in Figure 6.57.

Ultimately, I do not see that application of both soft drop grooming and selection cuts to tagged b-jets improves seperation between $t\bar{t}$ and $t\bar{t}H$ events.

6.11 Truth Jets vs Reconstructed Jets

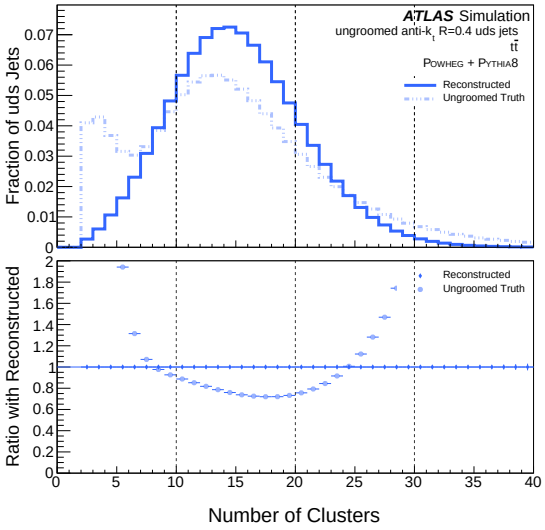
Following the investigations I make into the effects of soft drop grooming on flavour-tagged small jets reconstructed from Calorimeter Clusters, I extend this to explore the effects of soft drop grooming on matched truth-particle jets. I aim to explore how good these reco-jets are as a proxy for the real physics of the event, and whether applying soft drop grooming can improve that. I look to how ungroomed and groomed reco and truth-particle jets agree, and examine the effects of soft drop grooming on observables the truth-particle jets themselves. I hope to see whether the application of soft drop grooming to these different types of jets has different effects.

6.11.1 Number of Clusters

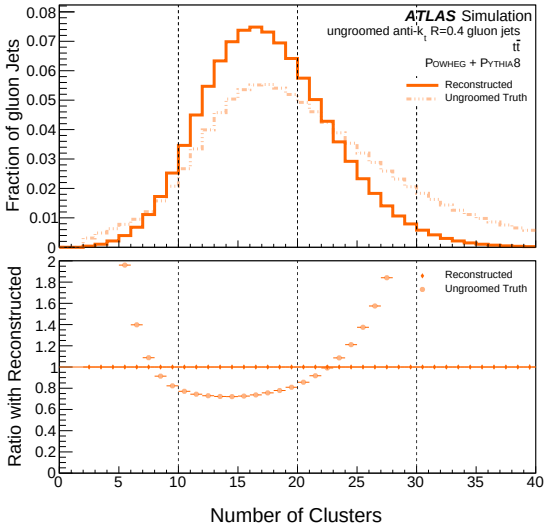
Firstly, I compare NCLUS for ungroomed reco and truth-particle jets, shown in Figure 6.58. As seen before, at an ungroomed level, flavour-tagged small jets reconstructed from calorimeter clusters have strongly similar distribution shapes, peaking between values of ten and sixteen. However looking at the results from the truth-jets I can determine broadly three groups. Firstly, gluon-, b- and c-jets, each with a very similar distribution shape in both truth and reco-jets. In each case, values for truth-jets cover a broader range than that of reco-jets. The peak value for gluon- and b-jets is a little higher in truth than reco, indicating some clusters are possibly missed in reconstruction. The peak for c-jets is slightly higher in reco than in truth, indicating that extraneous clusters could be wrongly included in reco-jets. Secondly, I inspect the distribution for truth uds-jets, which has a first small peak at a NCLUS value of five, followed by a second peak at thirteen, which is close to the peak for reco-jets, fourteen. Finally, I can look at the distributions for truth tau-jets, which is about half the width of the distribution of NCLUS for reco-jets with a peak in truth at close to half the value of the peak in reco tau-jets. This indicates many extraneous clusters are included in the reco-jets that are not present in truth tau-jets.

I expand this assessment by comparing the value of NCLUS for each given reco-jet against that of the truth-jet it is matched to, shown in Figure 6.59. For the quark- and gluon-jets I can see a broader distribution in truth-jets NCLUS and the shift in peaks. For uds-jets I also see the presence of the smaller first peak in the distribution of truth-jet NCLUS, I see that these truth-jets of two to three clusters are most strongly matched to reco-jets with between six and ten clusters. Looking at the results for tau-jets I see the broader distribution for reco when compared to truth-jets.

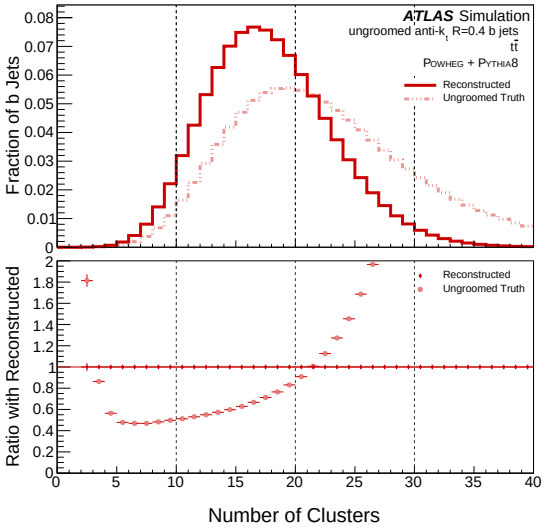
Firstly, I can compare values of NCLUS in ungroomed truth-jets to reco-jets that have been groomed with a soft drop grooming strength of $Z_{cut} = 0.05$, shown in Figure 6.60. Firstly I can see that for each of the quark-jet flavours the effect of applying soft drop grooming to reco-jets



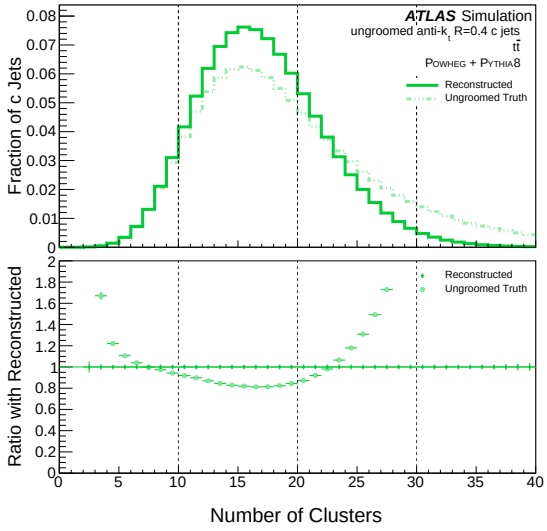
(a) uds



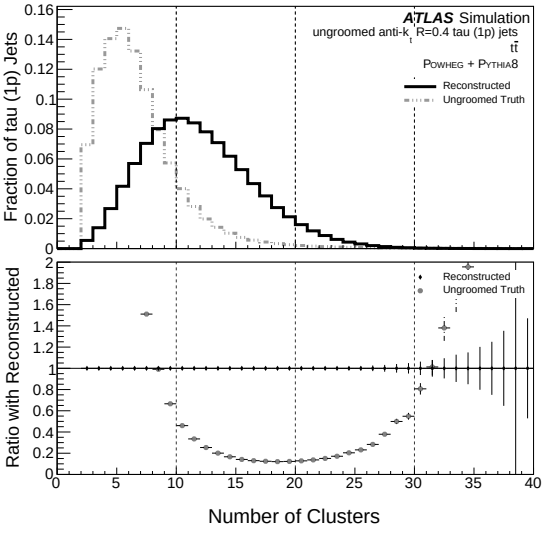
(b) gluon



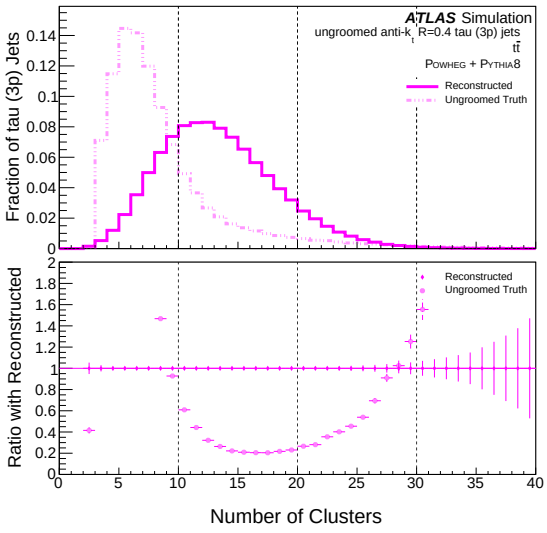
(c) b



(d) c

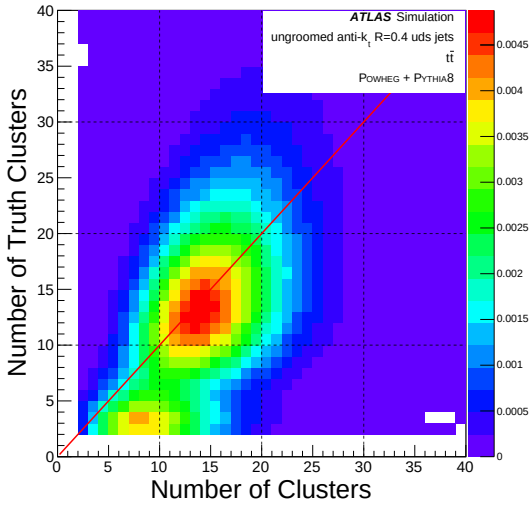


(e) 1-pronged tau

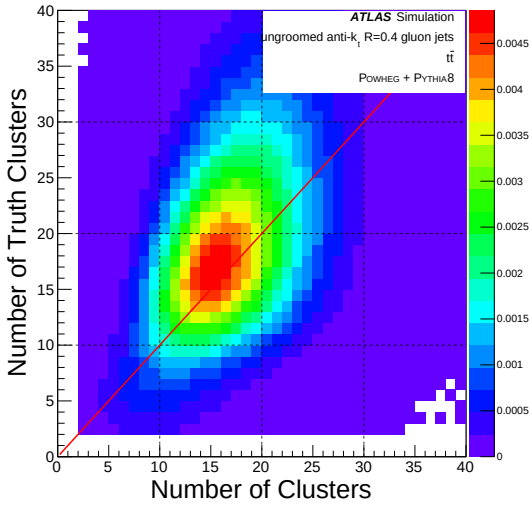


(f) 3-pronged tau

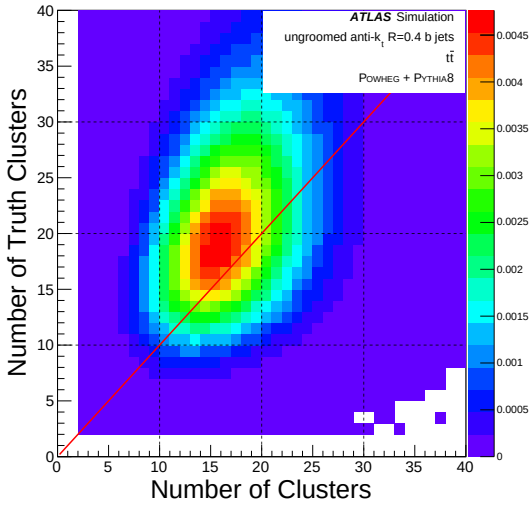
Figure 6.58: NCLUS in ungroomed reco and truth-particle jets.



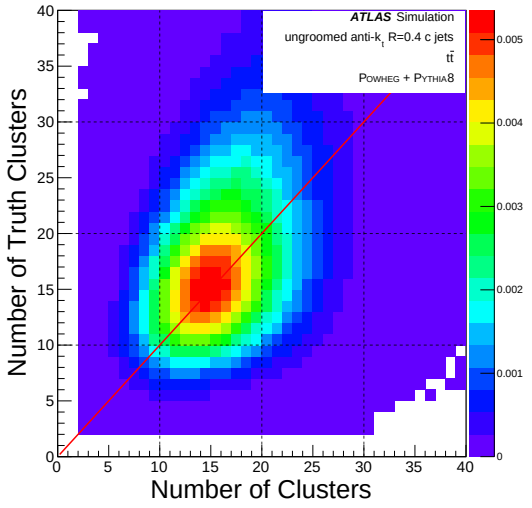
(a) uds



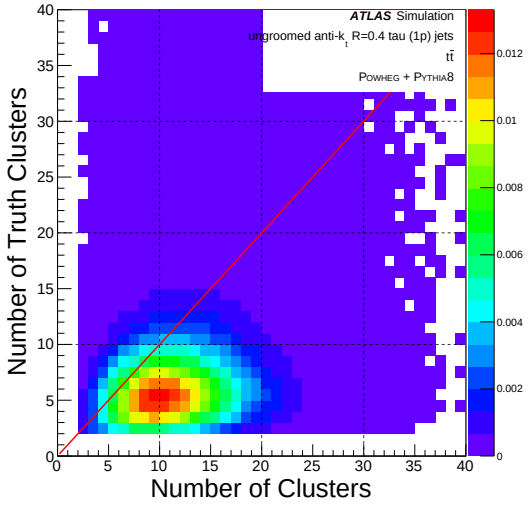
(b) gluon



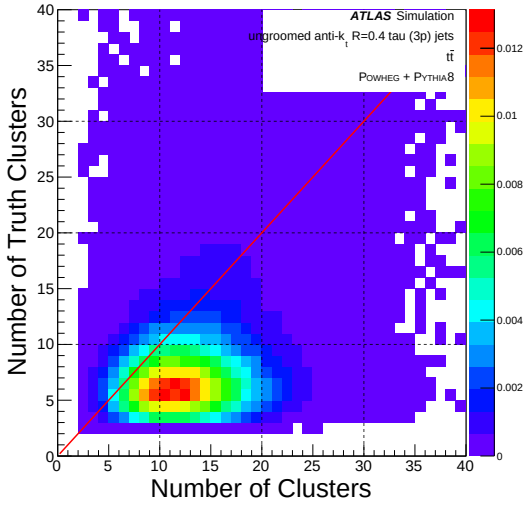
(c) *b*



(d) *c*

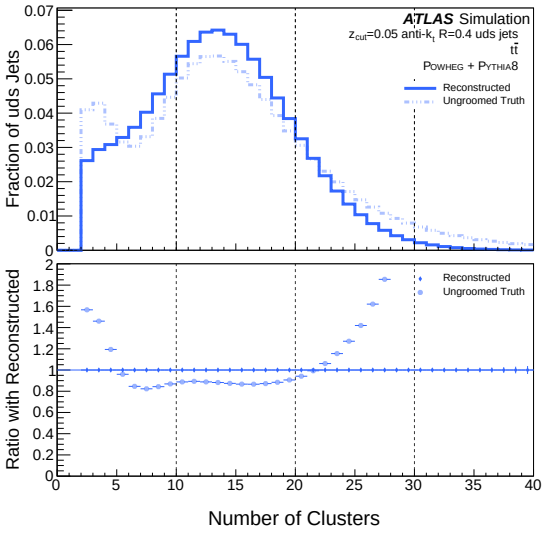


(e) 1-pronged tau

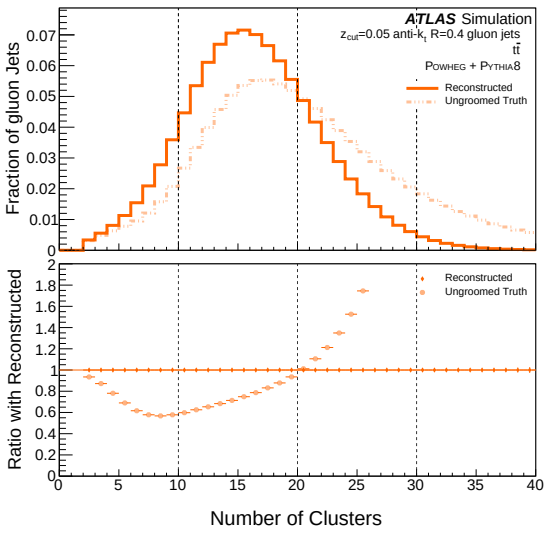


(f) 3-pronged tau

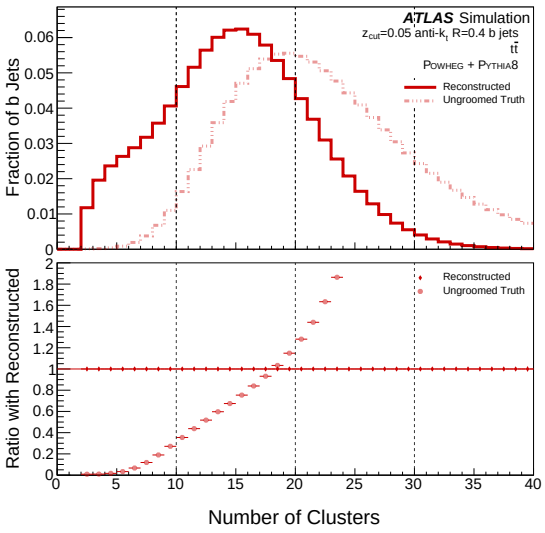
Figure 6.59: NCLUS for ungroomed matched reco and truth-particle jets.



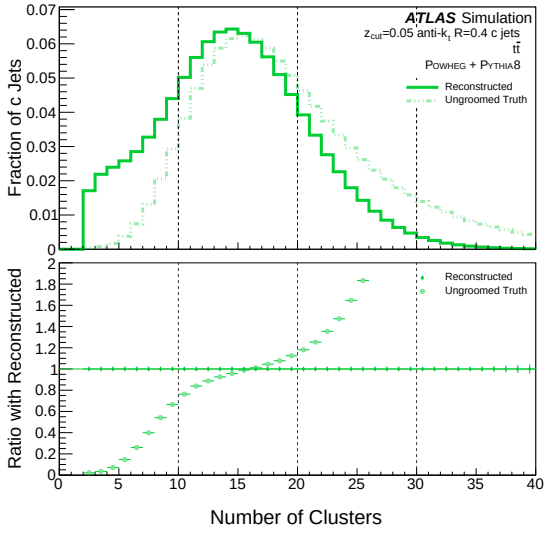
(a) uds



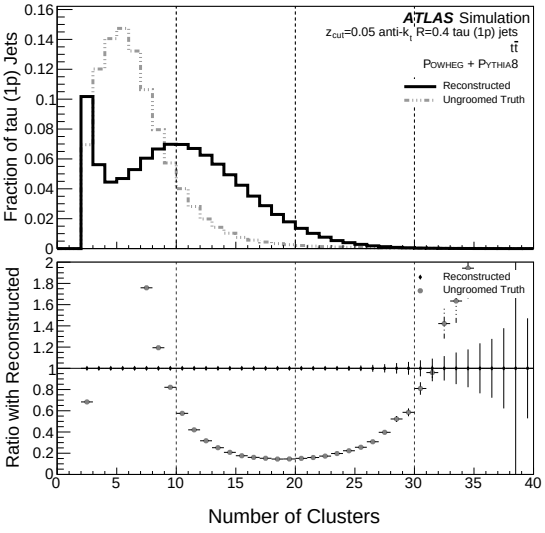
(b) gluon



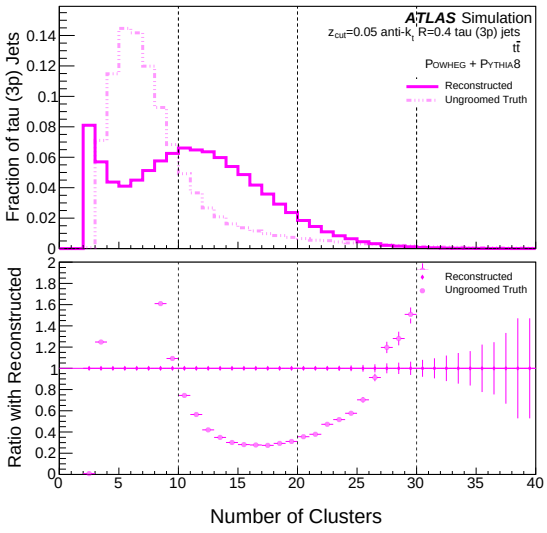
(c) *b*



(d) *c*

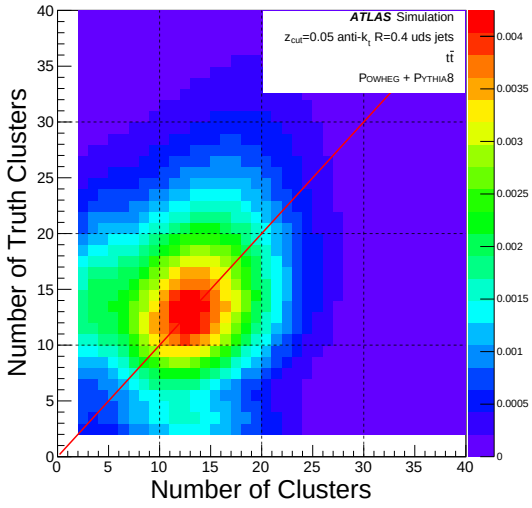


(e) 1-pronged tau

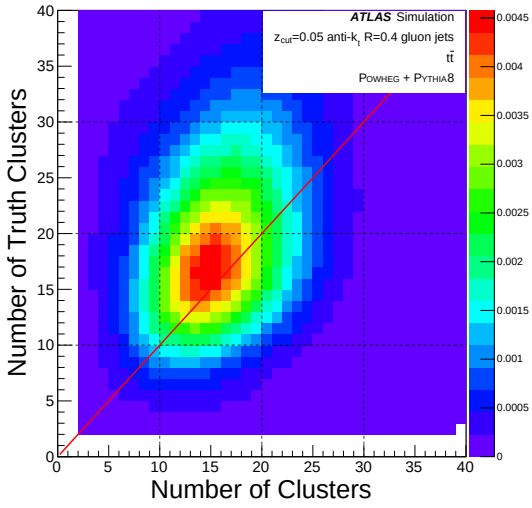


(f) 3-pronged tau

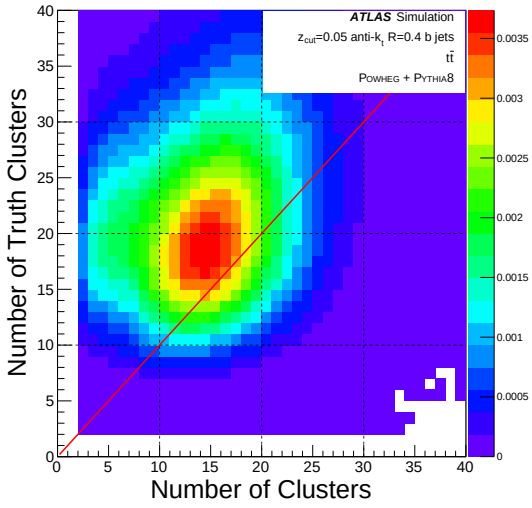
Figure 6.60: NCLUS in reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.05$ applied to reco-jets.



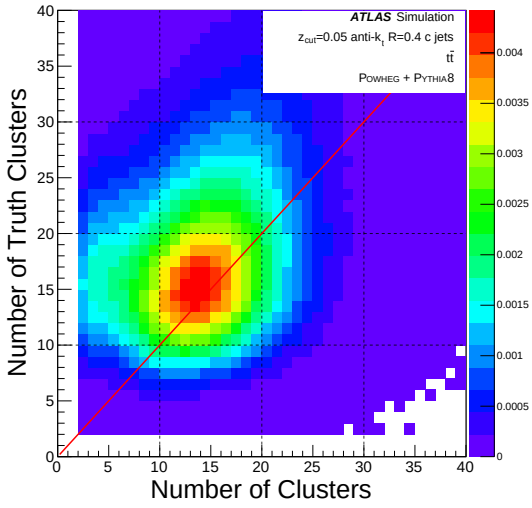
(a) uds



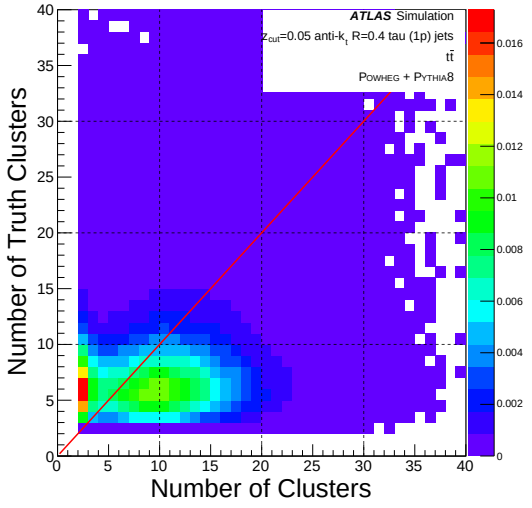
(b) gluon



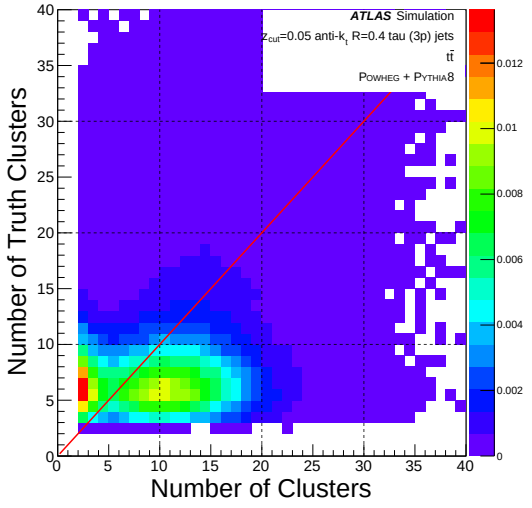
(c) *b*



(d) *c*

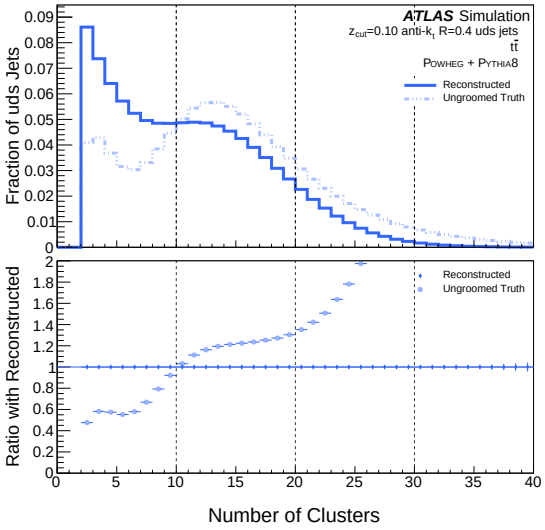


(e) 1-pronged tau

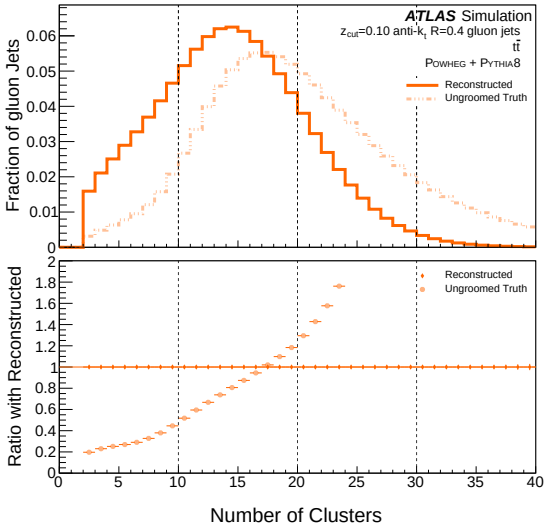


(f) 3-pronged tau

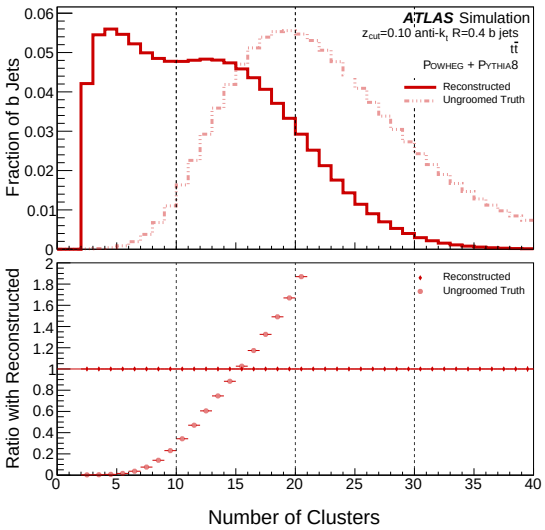
Figure 6.61: NCLUS for matched reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.05$ applied to reco-jets.



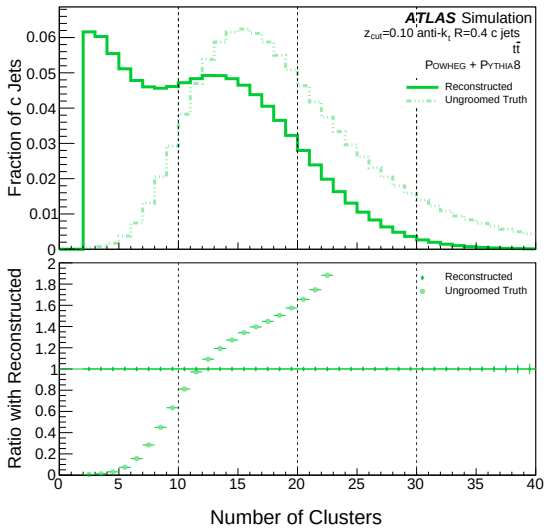
(a) uds



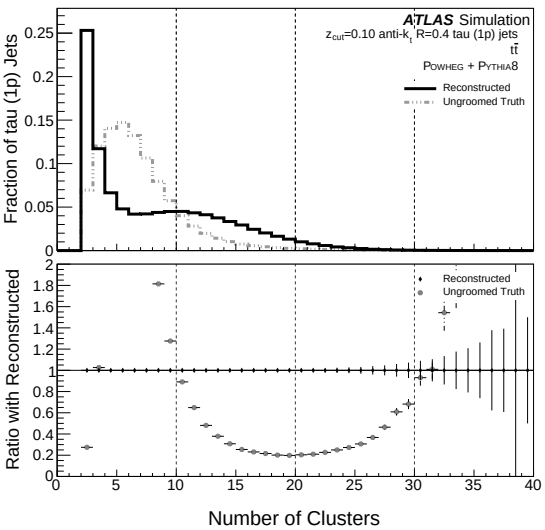
(b) gluon



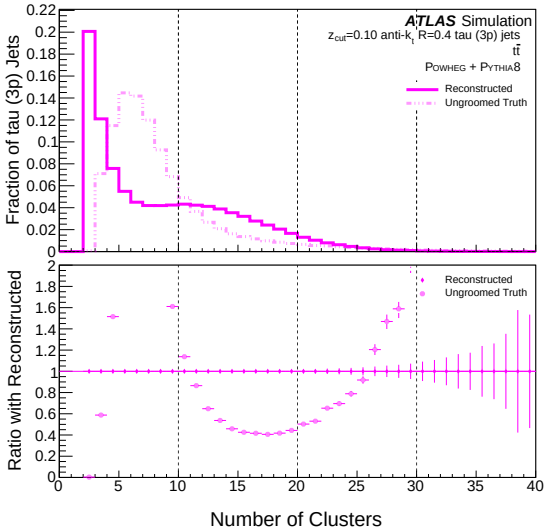
(c) *b*



(d) *c*

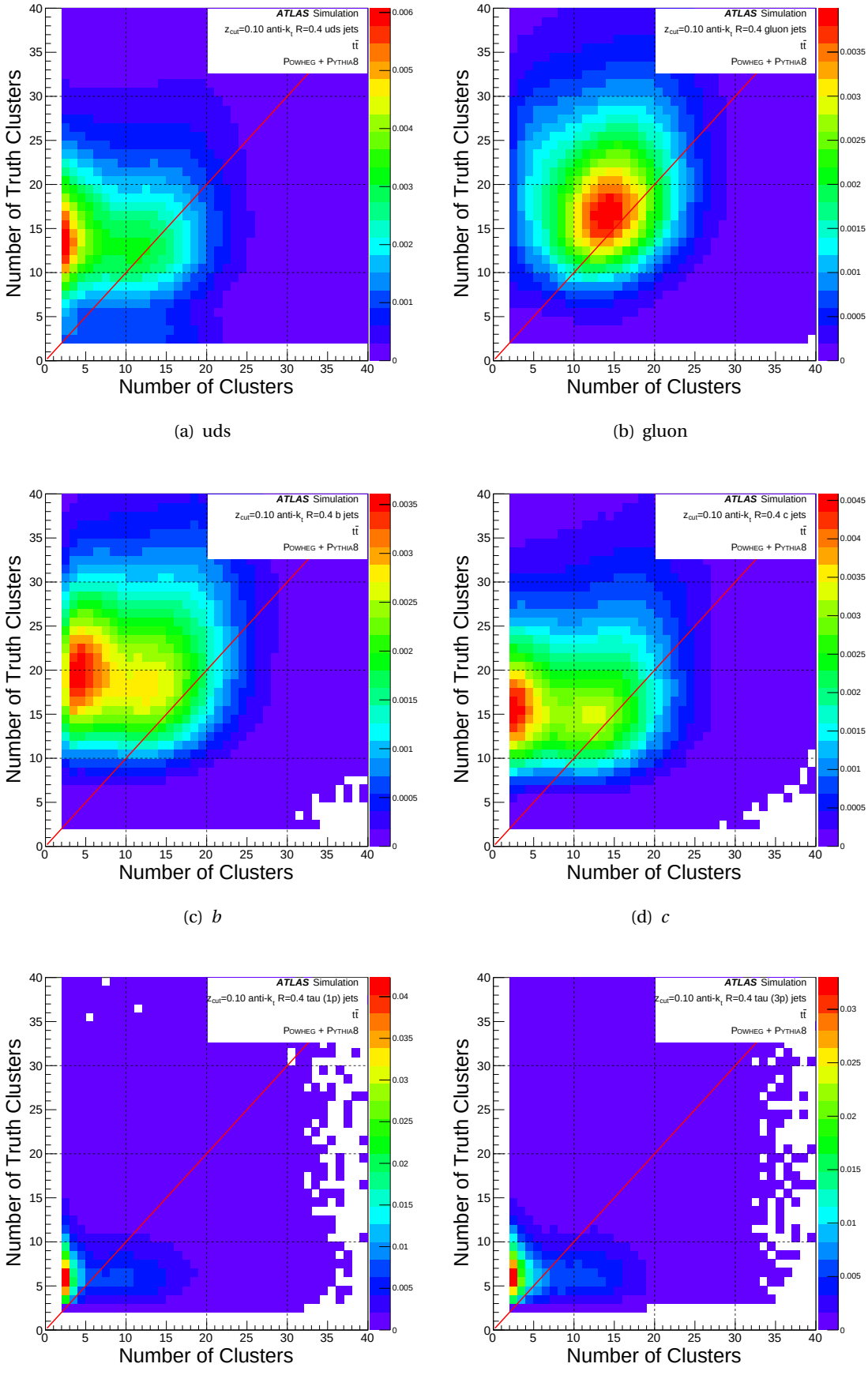


(e) 1-pronged tau



(f) 3-pronged tau

Figure 6.62: NCLUS in reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.1$ applied to reco-jets.

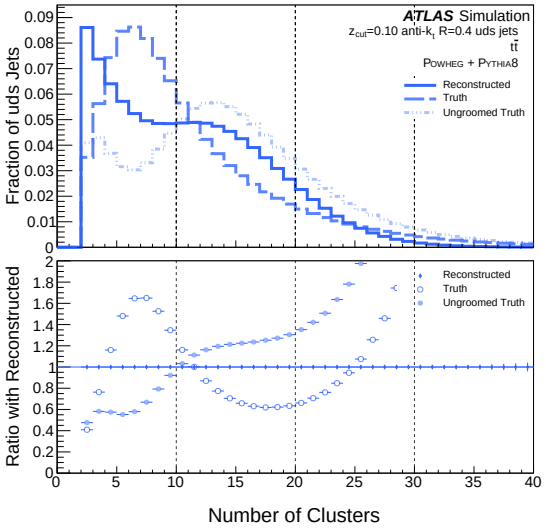


(a) uds (b) gluon

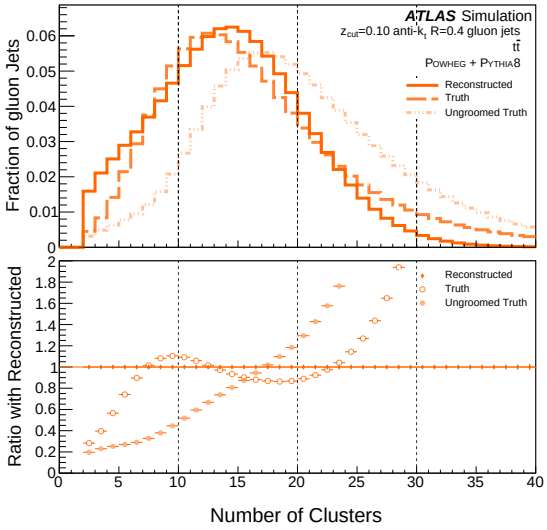
(c) b (d) c

(e) 1-pronged tau (f) 3-pronged tau

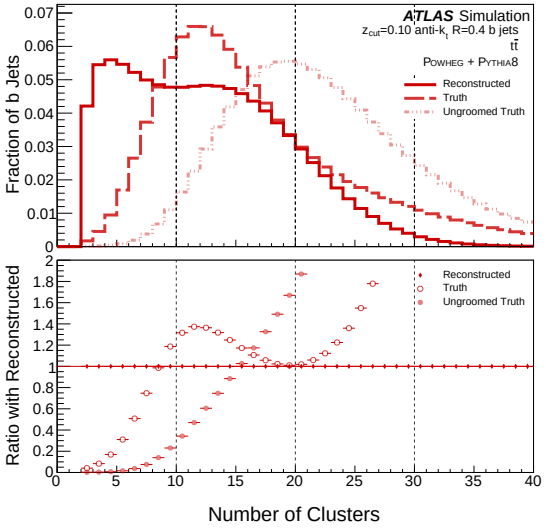
Figure 6.63: NCLUS for matched reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.1$ applied to reco-jets.



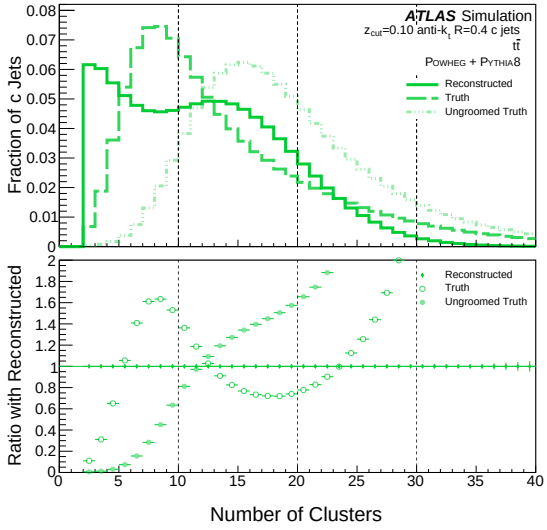
(a)



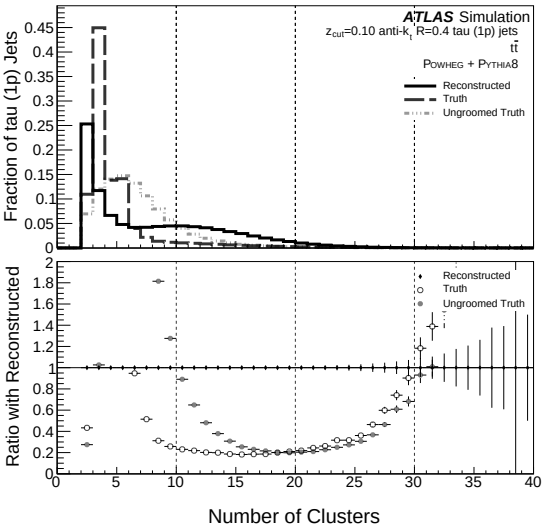
(b)



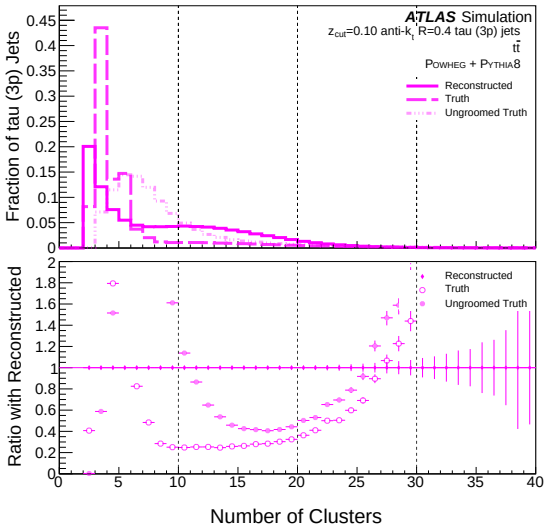
(c)



(d)



(e)



(f)

Figure 6.64: NCLUS in reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.1$ applied to both reco and truth-particle jets.

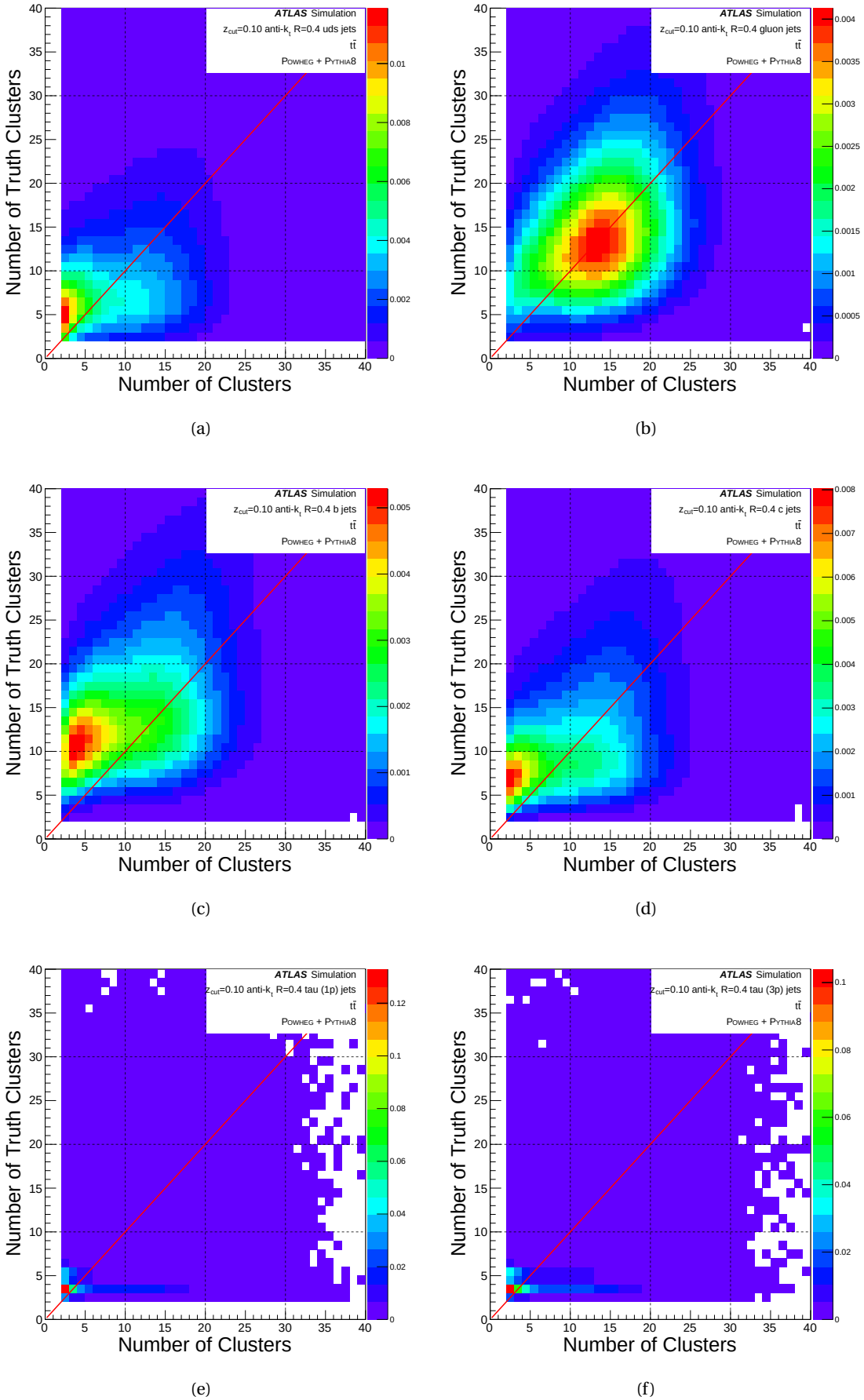


Figure 6.65: NCLUS for matched reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.1$ applied to both reco and truth-particle jets.

is that the distribution shape is now closer to that of the truth-jets. We see this in uds-jets, however the reco distribution is missing the first, low-value peak present in truth-jets. Both b- and c- reco-jet distributions become broader, with lower peaks, and therefore closer to the shape of their respective truth-jet distributions, however the peak value for reco b-jets moves further away from that of the truth-jets following the grooming away of clusters. Following the application of soft drop grooming strength of $Z_{cut} = 0.05$, the peak value of reco gluon-jets also moves to a lower value, and is subsequently further away from that of the truth-jets, however the distribution shape does not change, indicating an even distribution in NCLUS loss amongst reco-jets. In tau-jets I start to see the development of a higher, first peak close to the minimum value of NCLUS of two, which is not present in truth-jets.

Once more I look to the jet-by-jet comparison for NCLUS in ungroomed truth-jets and reco-jets with soft drop grooming applied, shown in Figure 6.61. In results for each quark flavour I start to see an increasing number of cases of reco-jets with five clusters or fewer matched to truth-jets with ten or more, demonstrating the effects of grooming away clusters within the reco-jets. In gluon-jets I see only a small change, with a uniform drop in NCLUS for reco-jets. In tau-jets the dramatic change I see matches the previous results, with the majority of matches showing two clusters in reco-jets. These are mostly matched to truth-jets with between four and seven clusters, indicating that low-p_T clusters both from the hadronic and from external sources, are groomed away.

Following this I can compare values of NCLUS in ungroomed truth-jets to reco-jets that have been groomed with a soft drop grooming strength of $Z_{cut} = 0.10$, shown in Figure 6.62. With this level of soft drop grooming applied I see the distributions for each of the sets of reco-jets changing to be more different to those of their respective truth-jets. For reco uds-jets I see the a more pronounced peak at the minimum value of NCLUS, 2. This is close to the location of the first peak in the truth-jet distribution, however it is much higher in reco-jets. This is followed by a supressed second peak in the reco distribution, close to that in the truth distribution, but slightly lower. The reco distribution for both b- and c-jets are similar, with high initial peaks at very low values of NCLUS, followed by a smaller, less noticeable peak that is closer to that in truth-jets, but smaller and of a lower value. In gluon-jets I see little difference from the previous grooming, but a slightly broadened reco distribution means it is closer to the truth distribution. In tau-jets the reco-jet distribution I see is no closer to that of truth-jets. Over all these reco-jet distributions seem to be slightly "over-groomed" for the observable NCLUS.

Looking to results for the matched jet comparison, in Figure 6.63, I see a continuation in the trend of previous results. The distributions for quark-jets show truth-jets with NCLUS values between ten and twenty-five matched to reco-jets with fewer than five clusters. The effect of the stronger strength of soft drop grooming is apparent here. In gluon-jets I start to see the effects of grooming reducing the number of clusters in reco-jets more clearly. In tau-jets I start to see very few reco-jets with more than five clusters.

We can also see how applying soft drop grooming of strength $Z_{cut} = 0.10$ to both truth and reco-jets comparably affects jet observables such as NCLUS. The results of this is shown in

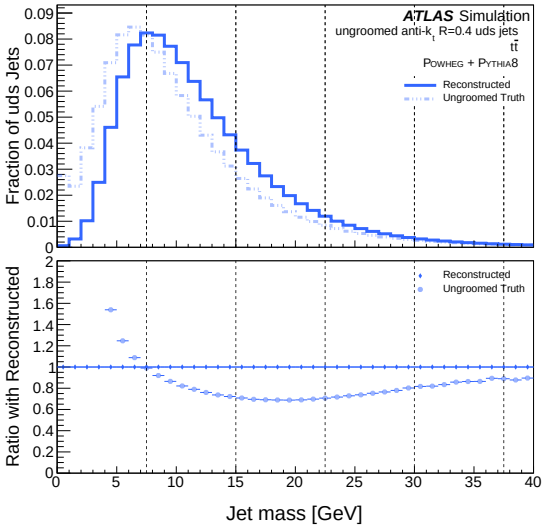
Figure 6.64. In the three quark-flavours I can see a similar effect soft drop grooming has on the NCLUS within truth-jets, in each case I see a simple, narrow distribution with peak at a lower value than for ungroomed truth-jets. The effect of soft drop grooming of strength $Z_{cut} = 0.10$ on NCLUS in truth-jets is similar to that of soft drop grooming strength $Z_{cut} = 0.05$ on reco-jets, and less severe than reco-jets groomed with strength $Z_{cut} = 0.10$– demonstrating these truth-jets have fewer soft, wide-angle clusters and are therefore less susceptible to the effects of soft drop grooming. For gluon-jets I see that truth and reco-jets groomed with a strength of $Z_{cut} = 0.10$ have a very similar value for NCLUS, with truth-jets having a slightly lower value. Despite the change in reco gluon-jets due to soft drop grooming only being small, it is now very close to the groomed truth-jet. The small difference between groomed reco and truth in gluon-jets illustrates the small amount of corruption in NCLUS coming from external sources following soft drop grooming. In truth tau-jets I see a dramatic change in NCLUS following soft drop grooming with the development of a very prominent peak displaying the majority of groomed truth-jets have only three or four clusters, a result that shows a strong resemblance to groomed reco-jets. This indicates that soft drop grooming is likely targeting the same clusters in both truth and reco-jets leaving only the hard-core for both.

We can explore these effects further by comparing soft drop grooming on matched jet pairs, the results of which are shown in Figure 6.65. We see that for uds-, *b*- and *c*-jets the effect on the two-dimensional distribution is very similar: the value of NCLUS for truth-jets is generally shifted to lower values, but it seems this is loosely relational to the value of NCLUS in matched reco-jets. Those truth-jets matched to reco-jets with fewer clusters are more likely to lose clusters themselves from grooming. This indicates that the effects that soft drop grooming is having on reco-jets is happening in the same way within these matched truth-jets – the same information is being groomed away. For tau-jets I also see this, but to a stronger degree – in the same way the soft drop grooming has more of an extreme effect on reco tau-jets. In gluon-jets I see a more even distribution in the change of NCLUS for truth-jets – less related to NCLUS values in the respective matched reco-jets. This is likely related to the smaller effect soft drop grooming had on reco gluon-jets. Ultimately these results indicate that the same changes are taking place in both reco and truth-jets following soft drop grooming, if perhaps to a slightly reduced degree.

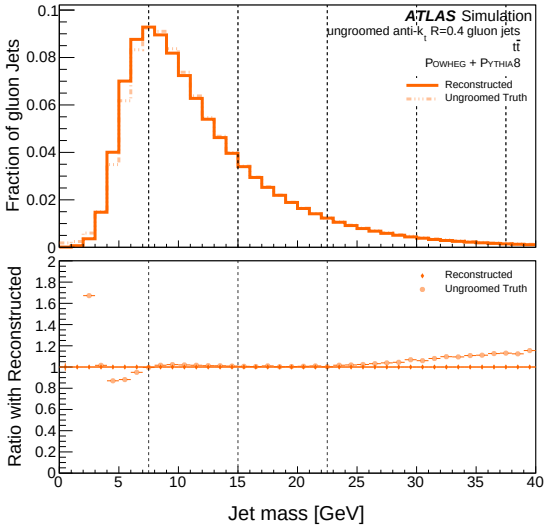
Overall these results show that applying soft drop grooming with a strength of $Z_{cut} = 0.05$ to reco-jets generally improves the agreement with truth-jets, but that applying strength of $Z_{cut} = 0.10$ might lead to "over-grooming" for values of NCLUS. This is not necessarily true for tau-jets which seem to have a more complicated relation. We also see the effects of applying soft drop grooming to truth-jets are comparable to their respective matched reco-jets, but to a lesser degree, indicating that they have fewer soft, wide-angle clusters that will be groomed away.

6.11.2 Jet Mass

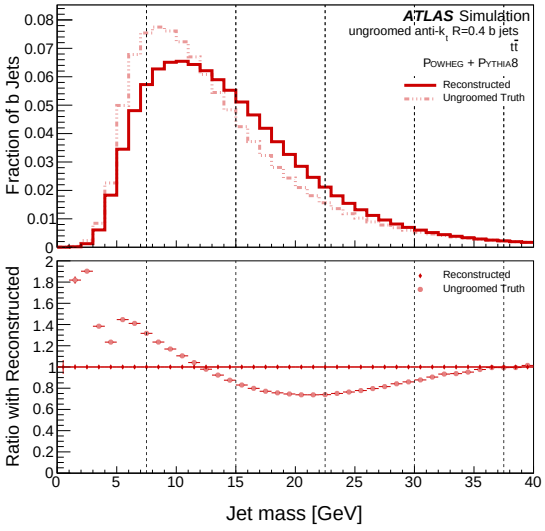
I start by contrasting the values of jet mass for ungroomed truth-particle and reco-jets, shown in Figure 6.66. I discern that reco gluon-jets show very close agreement to those in truth-



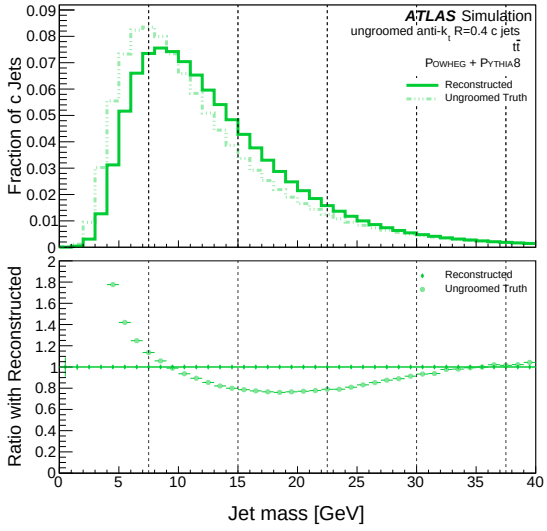
(a) uds



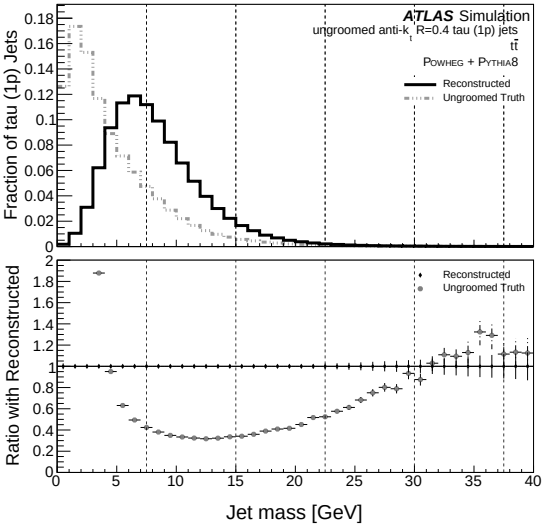
(b) gluon



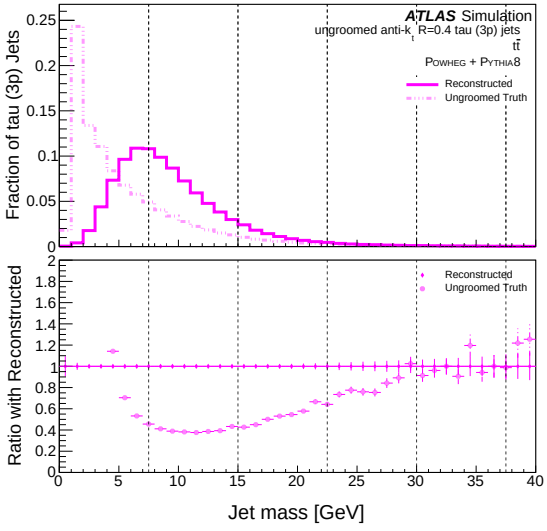
(c) b



(d) c

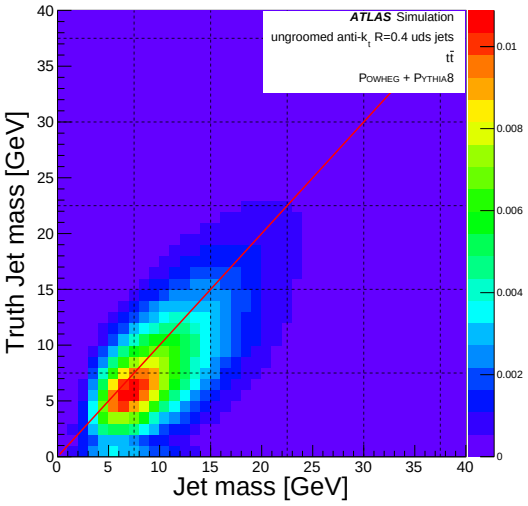


(e) 1-pronged tau

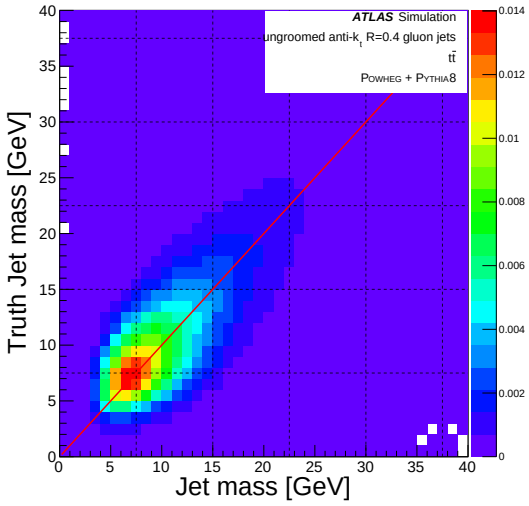


(f) 3-pronged tau

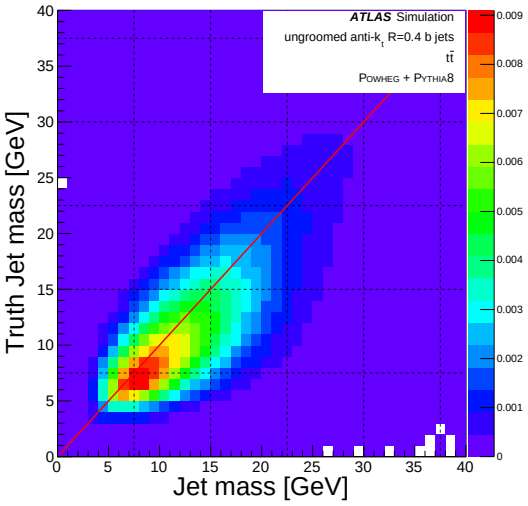
Figure 6.66: Jet mass in ungroomed reco and truth-particle jets.



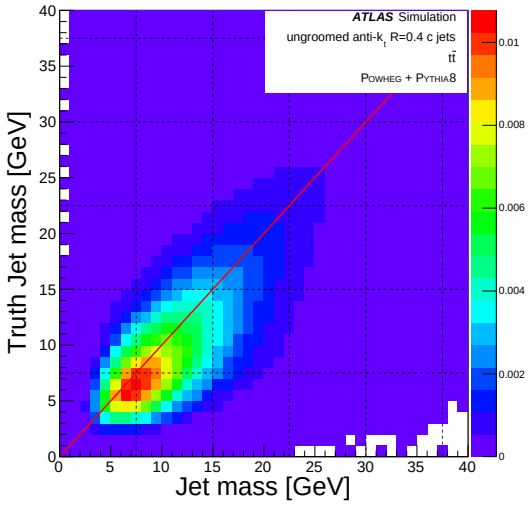
(a) uds



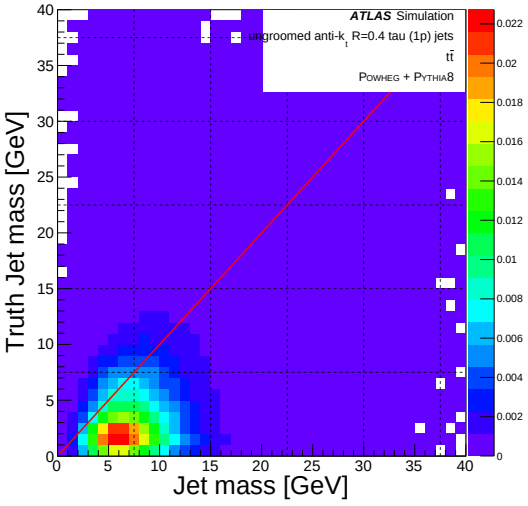
(b) gluon



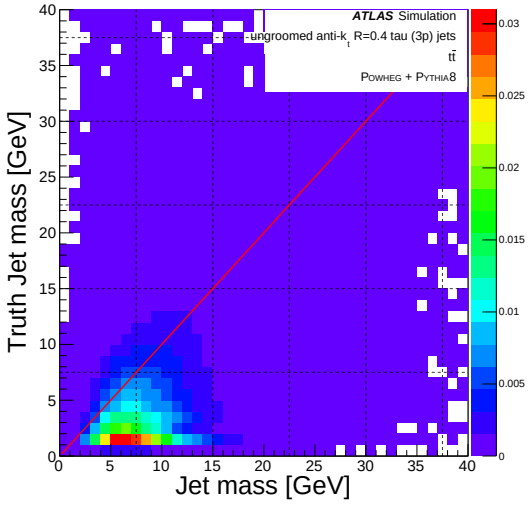
(c) *b*



(d) *c*

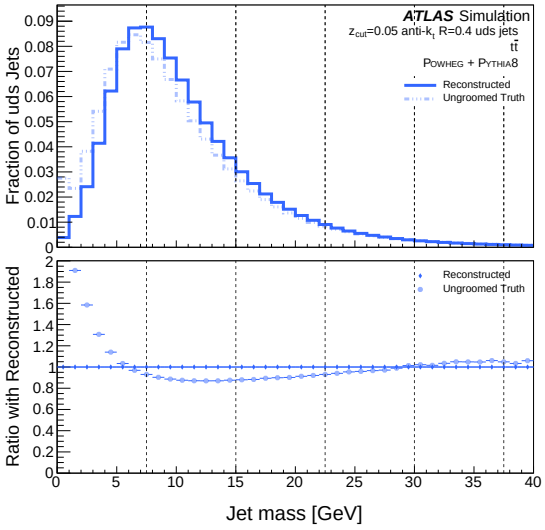


(e) 1-pronged tau

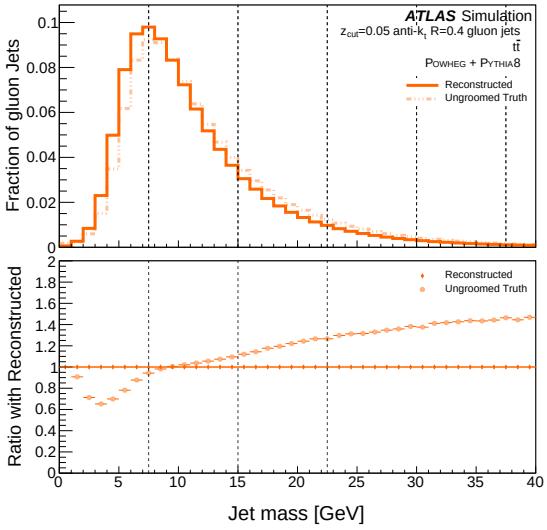


(f) 3-pronged tau

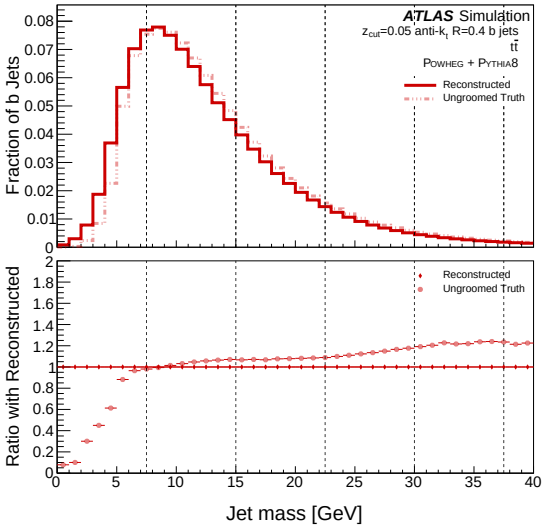
Figure 6.67: Jet mass for ungroomed matched reco and truth-particle jets.



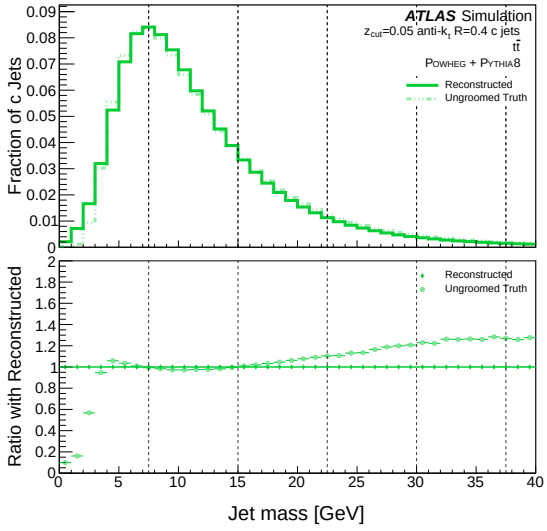
(a) uds



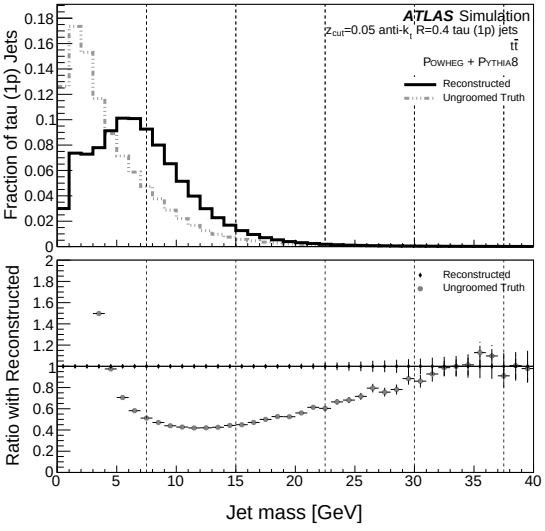
(b) gluon



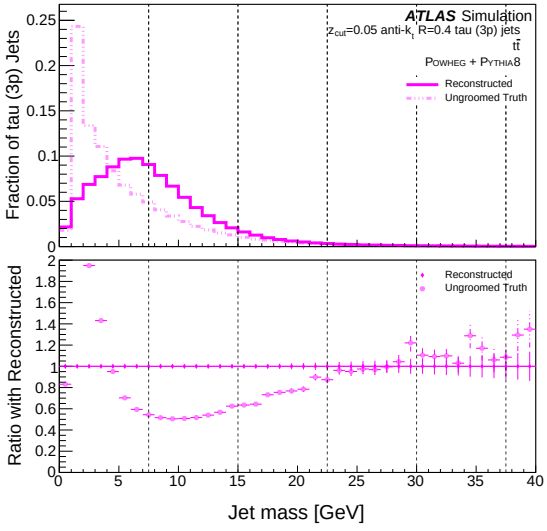
(c) *b*



(d) *c*

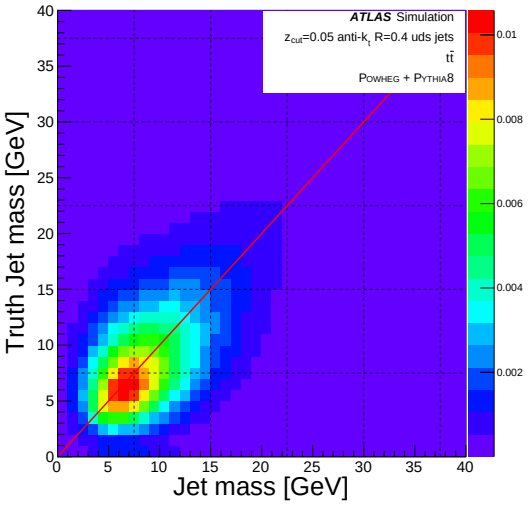


(e) 1-pronged tau

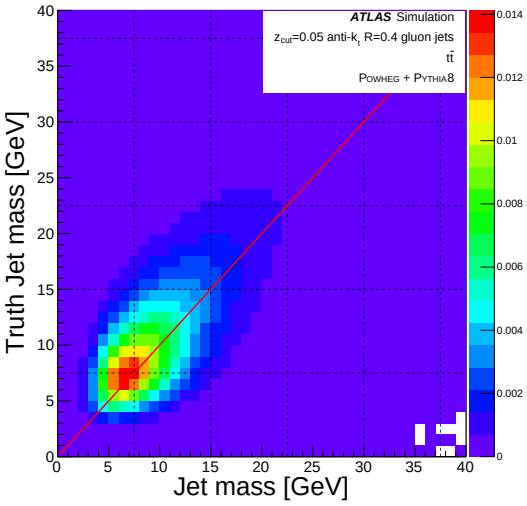


(f) 3-pronged tau

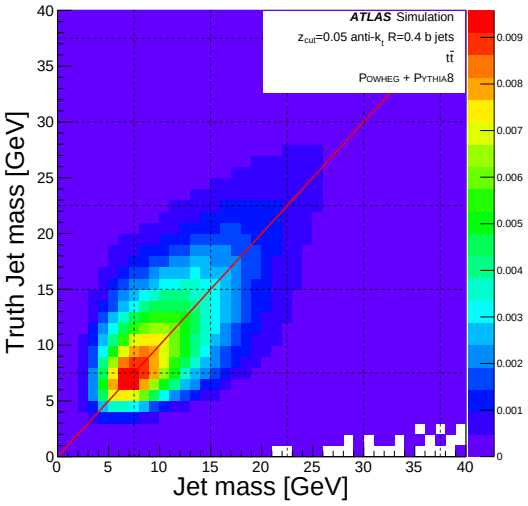
Figure 6.68: Jet mass in reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.05$ applied to reco-jets.



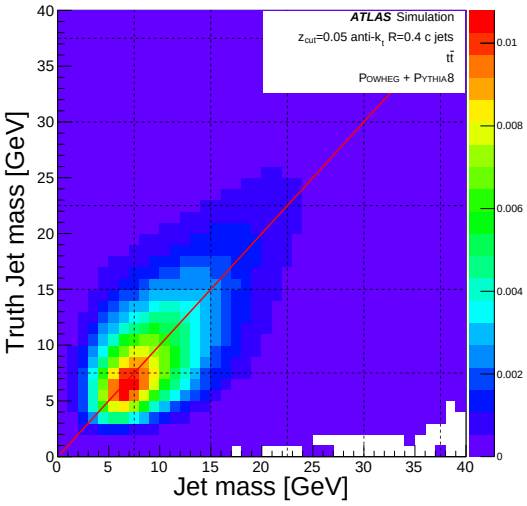
(a) uds



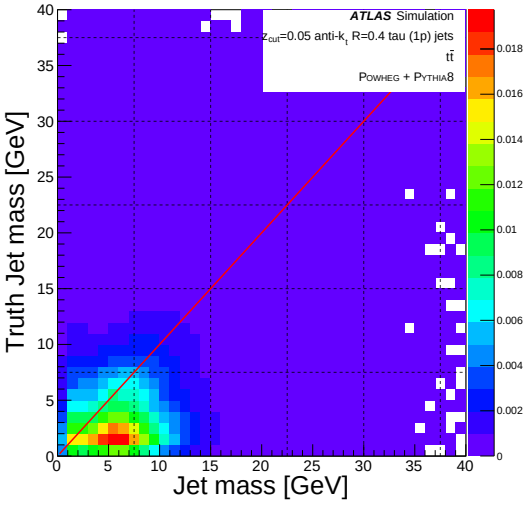
(b) gluon



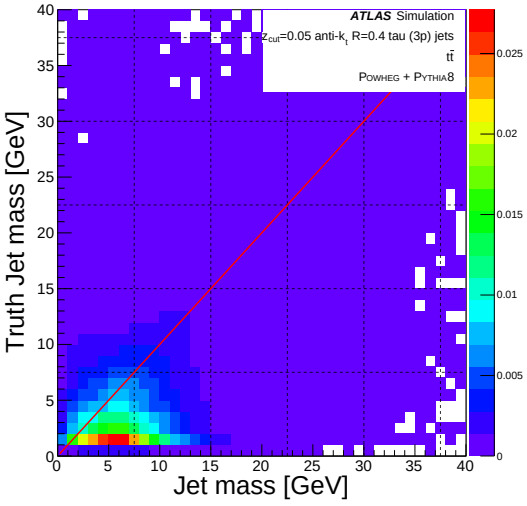
(c) *b*



(d) *c*

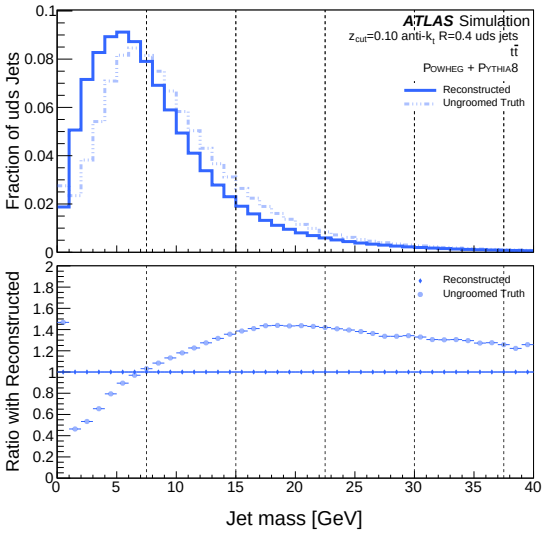


(e) 1-pronged tau

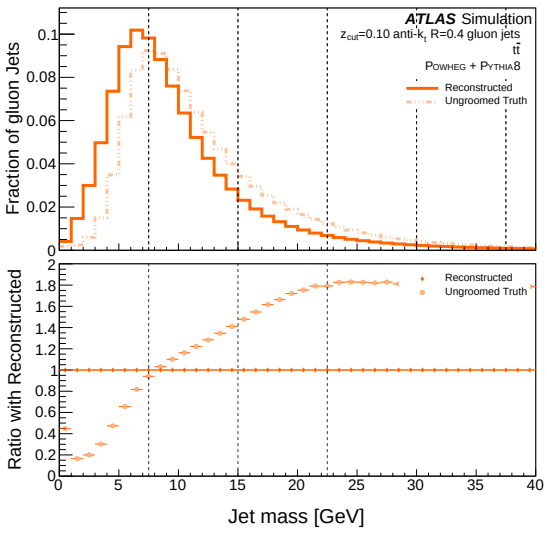


(f) 3-pronged tau

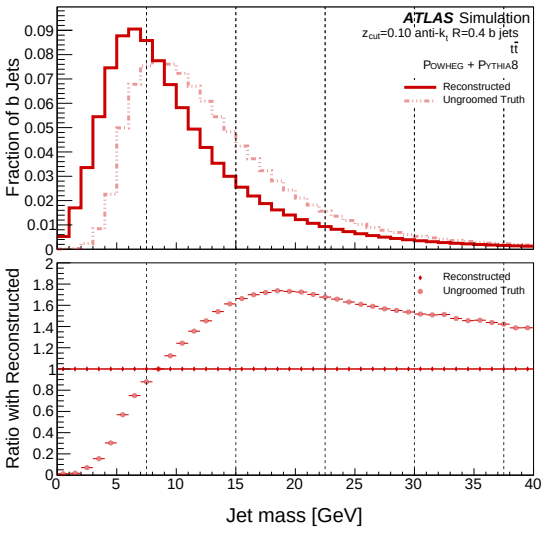
Figure 6.69: Jet mass for matched reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.05$ applied to reco-jets.



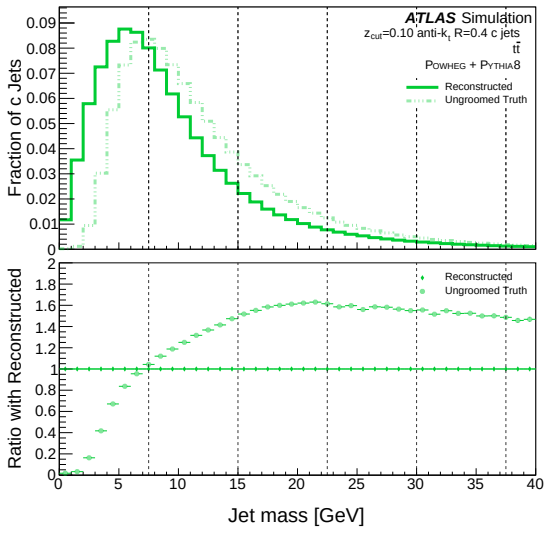
(a) uds



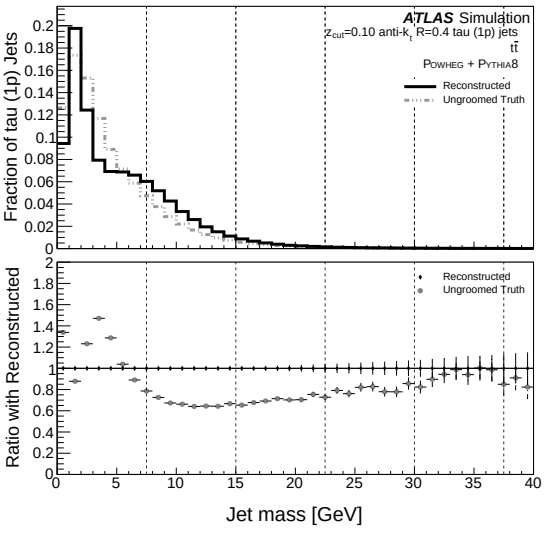
(b) gluon



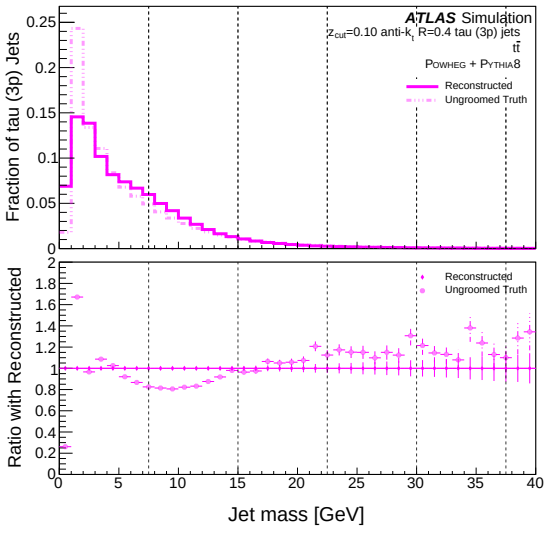
(c) *b*



(d) *c*

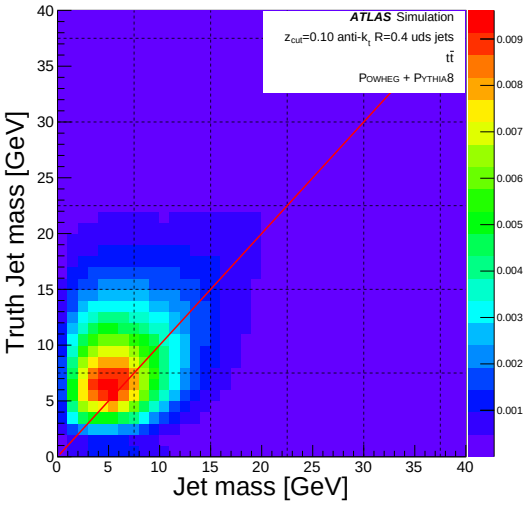


(e) 1-pronged tau

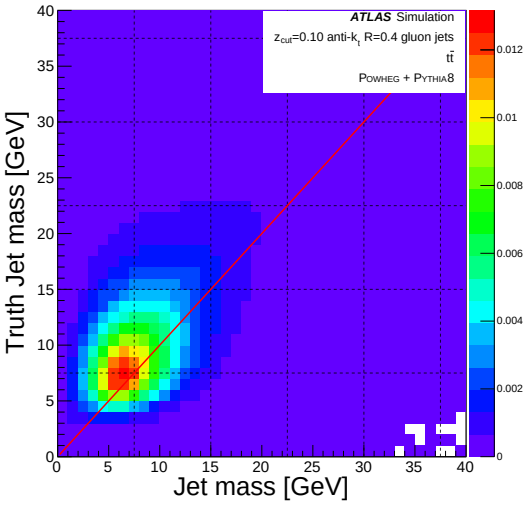


(f) 3-pronged tau

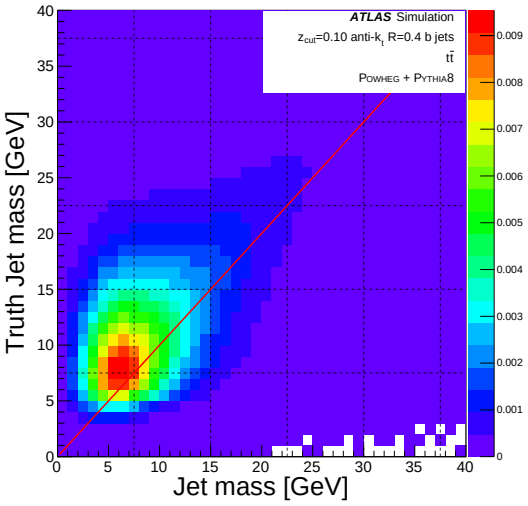
Figure 6.70: Jet mass in reco and truth-particle jets with soft drop grooming strength Z $_{\text{cut}}$ = 0.1 applied to reco-jets.



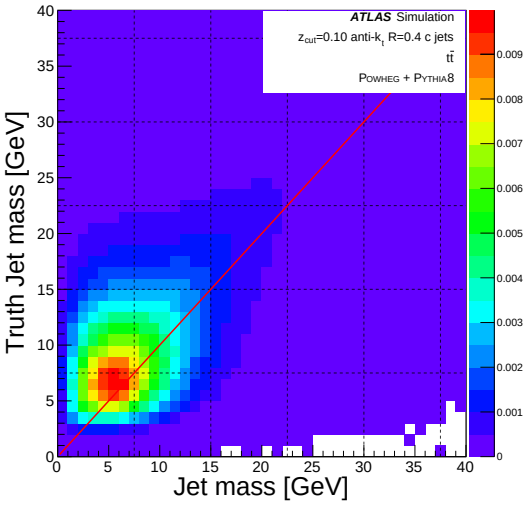
(a) uds



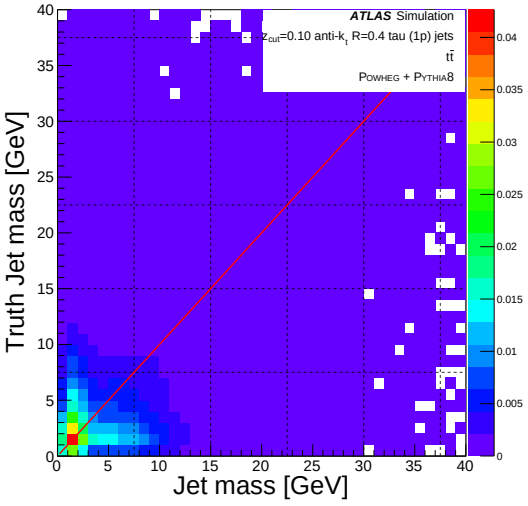
(b) gluon



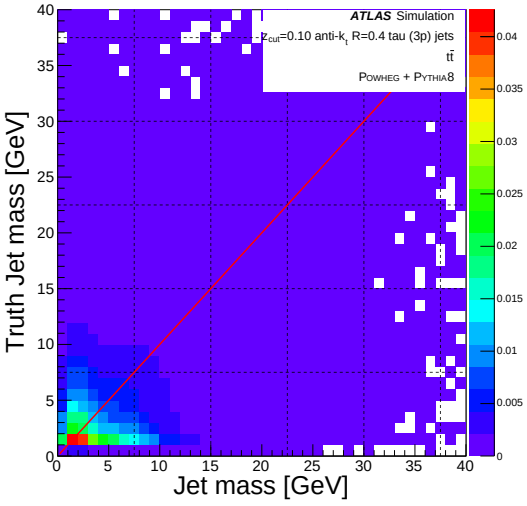
(c) *b*



(d) *c*

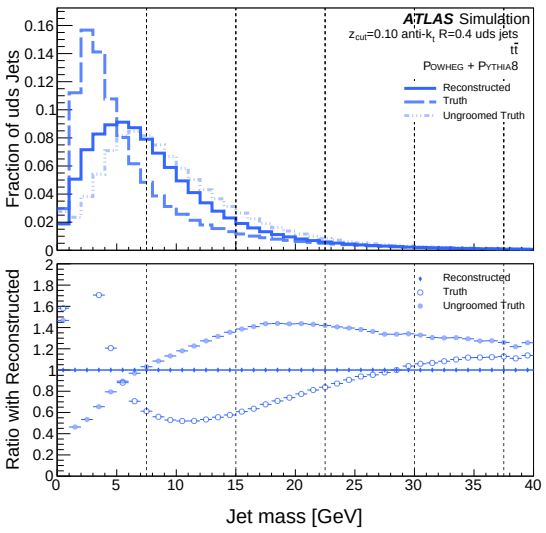


(e) 1-pronged tau

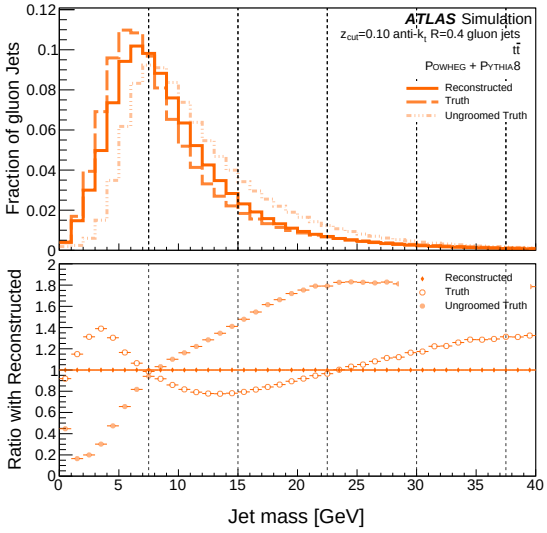


(f) 3-pronged tau

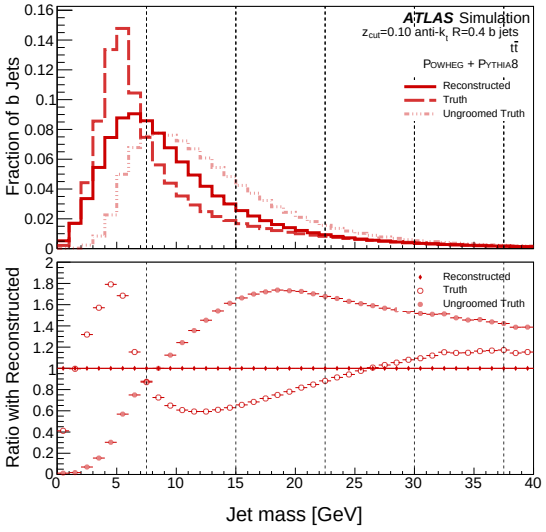
Figure 6.71: Jet mass for matched reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.1$ applied to reco-jets.



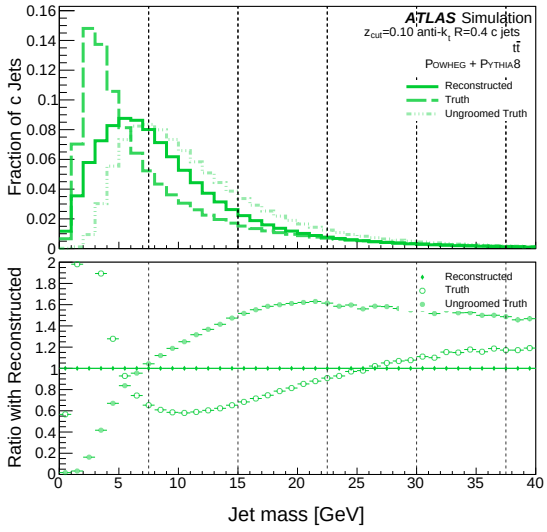
(a)



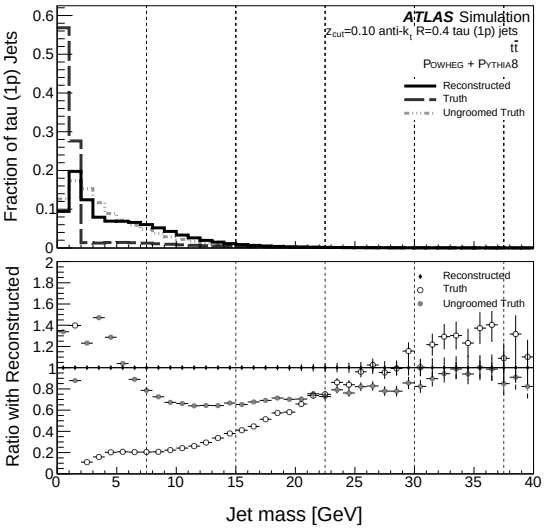
(b)



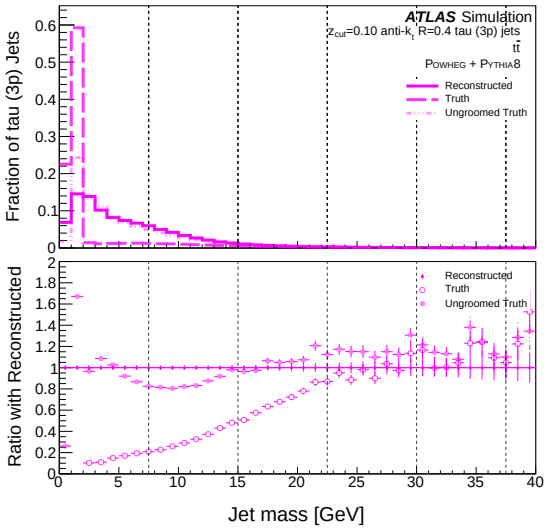
(c)



(d)

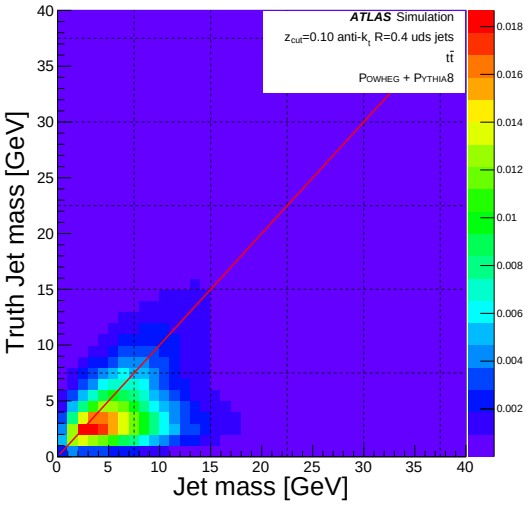


(e)

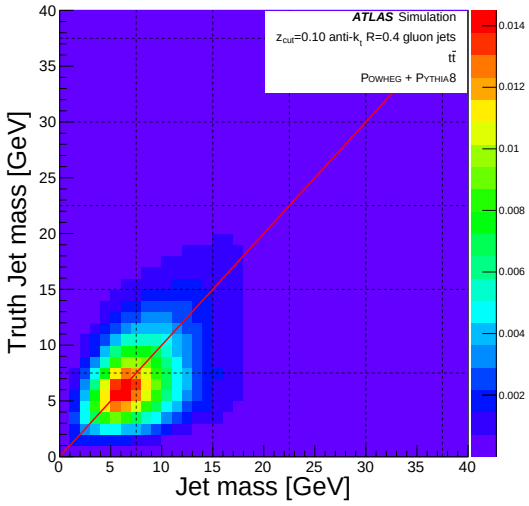


(f)

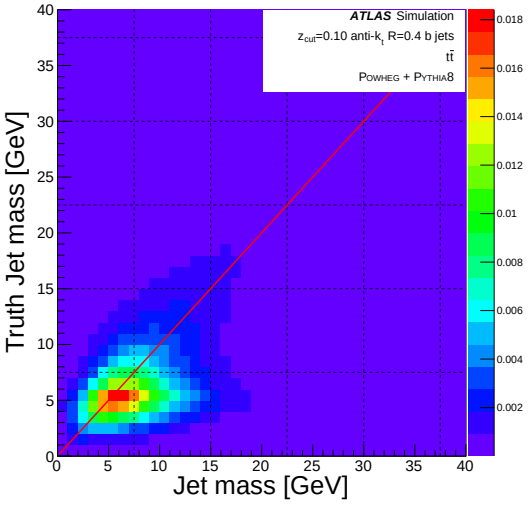
Figure 6.72: Jet mass in reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.1$ applied to both reco and truth-particle jets.



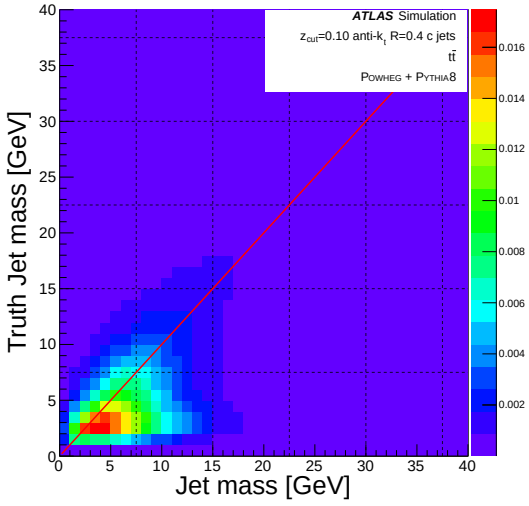
(a)



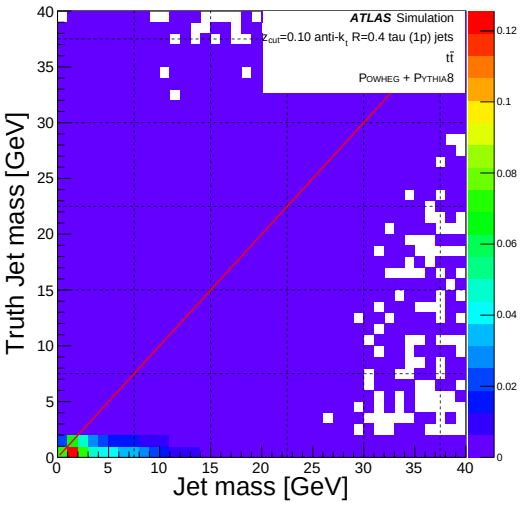
(b)



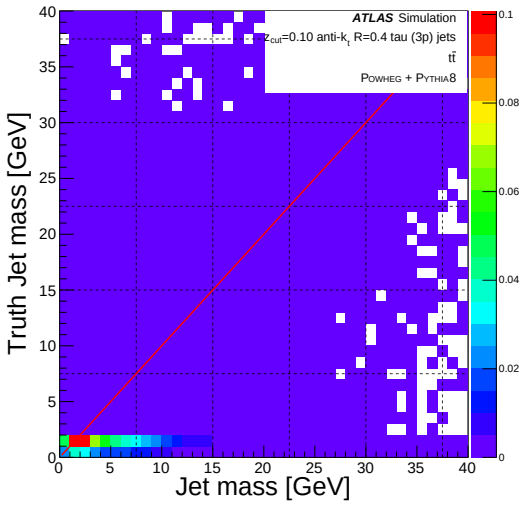
(c)



(d)



(e)



(f)

Figure 6.73: Jet mass for matched reco and truth-particle jets with soft drop grooming strength $Z_{cut} = 0.1$ applied to both reco and truth-particle jets.

particle gluon-jets when looking at jet mass. I see that all quark-jet flavours have similar results to one another: reco-jets have a slightly broader distribution and peak at a higher value of jet mass, when compared to their respective sets of truth-particle jets. This indicates the inclusion of small amount of mass from sources external to the truth-particle event. Results for tau-jets show very narrow truth-particle jets compared to non-tau-jets flavours. As the reco shape of tau-jets is similar to non-tau-jets flavours, this means the reco distribution is therefore much broader and higher peaking than the truth. This initially indicates the presence of lots of extra mass coming from external origins. I combine this information with that seen in the values of NCLUS. In gluon-jets the drop in jet mass from truth-particle to reco is smaller than the drop in NCLUS, this shows that the clusters missed in reco-jets were seemingly low mass. In uds- and c-jets the opposite is true. A small increase in NCLUS from truth-particle to reco and a larger increase in jet mass indicates the extra clusters included in reconstruction are higher mass. In b-jets the picture is more complicated with reco-jets having a lower value of NCLUS with a slightly higher jet mass compared to truth-particle jets – this could indicate a combination of missing low-mass clusters and incorporating fewer, high-mass clusters from external origins. In tau-jets see both the value of NCLUS and jet mass for reco-jets are essentially double that in truth, indicating only that the average jet mass of these extra clusters is comparable to those within the truth-particle jet.

Investigating these results on a jet-by-jet basis to compare truth-particle and reco results for jet mass can tell me more. I present these results in Figure 6.67. In gluon-jets I see a very strong correlation between the jet mass of a reco-jet and its respective matched truth-particle jet, once more demonstrating reco gluon-jets are a good proxy for truth-particle jets when looking at jet mass. I also observe a strong correlation for the quark flavours with only a slight skew towards higher values of jet mass in reco-jets. This indicates the extra jet mass seen in the distribution for reco-jets is likely evenly distributed amongst quark-jets. In tau-jets I percieve the replication of the previous result, with jet mass roughly double the value in reco-jets as in truth. This seems to be slightly more present in low mass jets, perhaps relating to the shift present in tau-jets with lower values of NCLUS.

We now move onto studying the effects of soft drop grooming on jet mass within the Truth and Reco Jets. Firstly, I can apply soft drop grooming of strength $Z_{cut} = 0.05$ to the reco-jets and compare these to the ungroomed truth-jets. We show the results for this in Figure 6.68. We now see a very strong agreement in values of jet mass between the quark flavours, particularly in c-jets. In gluon-jets I see that following soft drop grooming the agreement between reco and truth-jets has worsened, with a decreasing of jet mass causing a shift in the distribution. This could be indicative of possible "over-grooming". In tau-jets I see a slightly improvement in agreement between truth and reco-jets, with jet mass values of reco-jets broadening to include lower values, closer to those of truth-jets. As a general trend I see that the larger values of NCLUS groomed away in reco-jets has lead to a smaller loss in jet mass, showing the clusters lost were low mass, as should be expected from soft drop grooming.

Comparing the jet mass of the groomed reco-jets to the jet mass of ungroomed truth-jets I

can see more information on the changes soft drop grooming brings, see Figure 6.69. We note that for quark-flavour jets I see a very strong correlation between jet mass for reco-jets with a soft drop grooming strength of $Z_{cut} = 0.05$ applied and ungroomed truth-jets, with uniform improvement across all truth jet mass values. This uniform change is also present in gluon-jets, however the change is slightly detrimental to the agreement between truth and reco values of jet mass. The results for tau-jets shows the most change, and I see more change in the value of reco jet mass for truth-jets of lower jet mass, meaning these low-mass jets are more susceptible to the effects of soft drop grooming.

We now groom the reco-jets with the default soft drop grooming strength $Z_{cut} = 0.10$, the results of which are shown in Figure 6.70. For uds-, gluon, *b*- and *c*-jets I now see a very similar result, whereby reco-jets have distributions narrower, and lower in jet mass, than their respective truth-jets. These reco distributions now appear "over-groomed" compared to truth. In each case this decrease and difference is less than for values of NCLUS, demonstrating that the clusters groomed away are low in mass. For tau-jets, however, I now start to see a stronger agreement between reco and truth values of jet mass than at lower strength, or no soft drop grooming. The distribution shapes are much more similar, with peaks in reco very close to those in truth. The dramatic change in NCLUS from significantly too high to too low, matched with change of jet mass from much too high to very comparable indicates the clusters lost were very likely to be "extra", perhaps from pile-up, and the smaller NCLUS left are the most hard parts of the jet containing the main mass of the tau-jets.

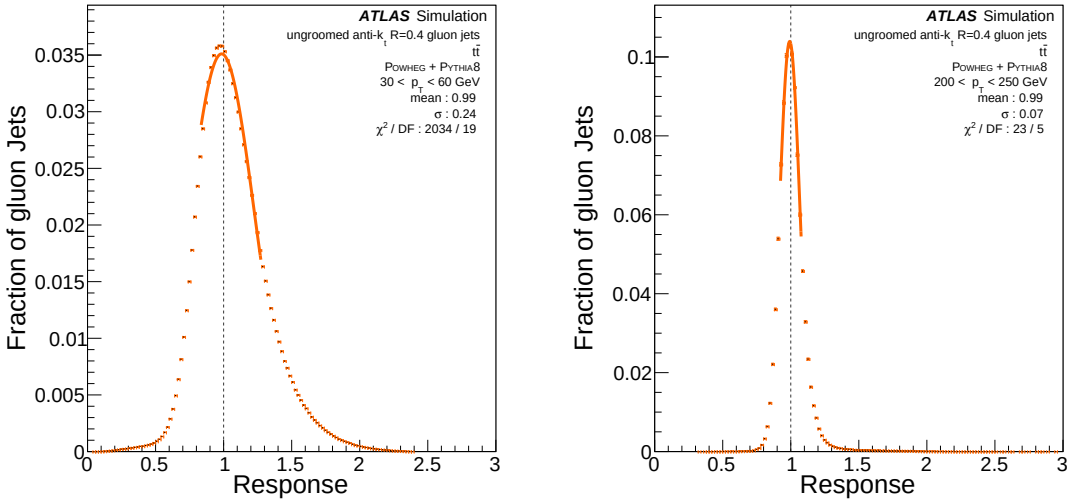
On a jet-by-jet basis the results of this strength of soft drop grooming is to slightly decrease the correlation between truth and reco jet mass for uds-, gluon,- *b*- and *c*-jets with a close uniform decrease in the value of jet mass. For tau-jets I see the obvious shift to lower values of jet mass in the reco-jets.

We can also explore the effects that soft drop grooming has on the truth-jets, shown in Figure 6.72. For every jet flavour I see that applying soft drop grooming with strength $Z_{cut} = 0.10$ has a bigger change on the distribution of jet mass in truth-jets than in reco-jets. In gluon-jets, those with the least change, I see that soft drop grooming has shifted the distribution of jet mass in truth-jets to slightly lower values than that of groomed reco-jets. In uds-, *b*- and *c*-jets I see that is has shifted much more with the development of a high peak at low jet mass. This is at 2 GeV for truth uds- and truth *c*-jets and at 5 GeV for truth *b*-jets, potentially presenting me with the hard-core of the truth-jet. For tau-jets the change is much more dramatic with over 80 % of each set of groomed truth tau-jets displaying jet mass between 0 GeV and 2 GeV. The change in jet mass for the truth-jets resulting from soft drop grooming appears to be more dramatic than changes in value of NCLUS for the same jets. This difference is opposite to that in reco-jets, indicating the clusters groomed away in truth-jets were much less likely to bewide-angle, low mass clusters, which I expect to see fewer of in the truth-jets as they originate most prominently from pile-up.

We can discover more about these changes in truth-jets by looking at matched comparisons for jet mass in truth and reco-jets. We see that in uds-, *b*-, *c*- and tau-jets that the drop in jet

mass within truth-jets does not appear to be totally uniform: the condensing of the distribution in truth mass direction seems to indicate more drop in value of jet mass from higher jet truth masses than lower. Ultimately the results from this section show me that reco-jets with different strengths of soft drop grooming applied, depending on the flavour, show strong agreement with truth-jets for values of jet mass: reco gluon-jets show strongest agreement when ungroomed; quark flavour jets, uds-, *b*- and *c*-jets, show strong agreement after a soft drop grooming of strength $Z_{cut} = 0.05$ has been applied; and reco tau-jets show good agreement when groomed with the chosen default soft drop grooming strength, $Z_{cut} = 0.$. We also note that applying soft drop grooming to truth-jets changes their jet mass more than when applied to reco-jets of their respective flavour – combining this with the smaller value of NCLUS generally lost by truth-jets from soft drop grooming illustrates the contrast to many more, lower-mass clusters groomed in reco-jets.

6.12 Jet Response



(a) p_T range 30 GeV to 60 GeV

(b) p_T range 200 GeV to 250 GeV

Figure 6.74: Jet response of ungroomed gluon-jets, in the p_T range 30 GeV to 60 GeV (b) and 200 GeV to 250 GeV (b)

In addition to exploring the consequences on JSS observables of applying soft drop grooming, I also investigate how it effects technical aspects of jets reconstruction in the ATLAS detector. A crucial element of this type is the *jet response* [201], a measurement that can tell me a lot about how much jet energy, or p_T is actually registered in the calorimeter during reconstruction. I therefore compare values of p_T in both the reco-jets and the truth-particle jet they are matched to, both before and after applying to soft drop grooming. I therefore define the measure of jet response by finding the p_T of the reco-jet as a fraction of the p_T of the truth-particle jet.

$$\frac{p_{T\,reco-jet}}{p_{T\,truth-particlejet}} \tag{6.2}$$

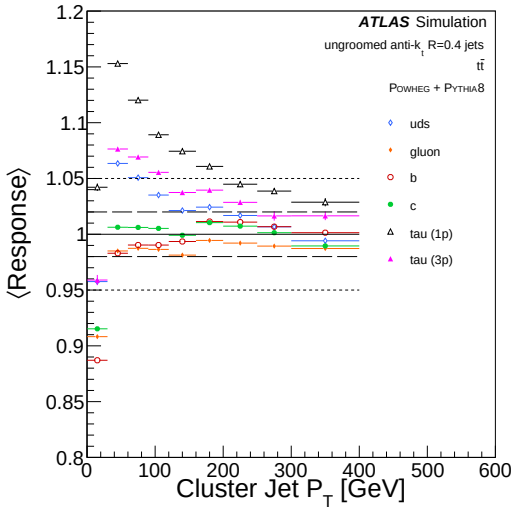


Figure 6.75: Mean jet response of all jet flavours, ungroomed.

For the analysis I look at this for each of the reco-jets, at each stage of grooming, and I always use the ungroomed truth-particle jet as the marker.

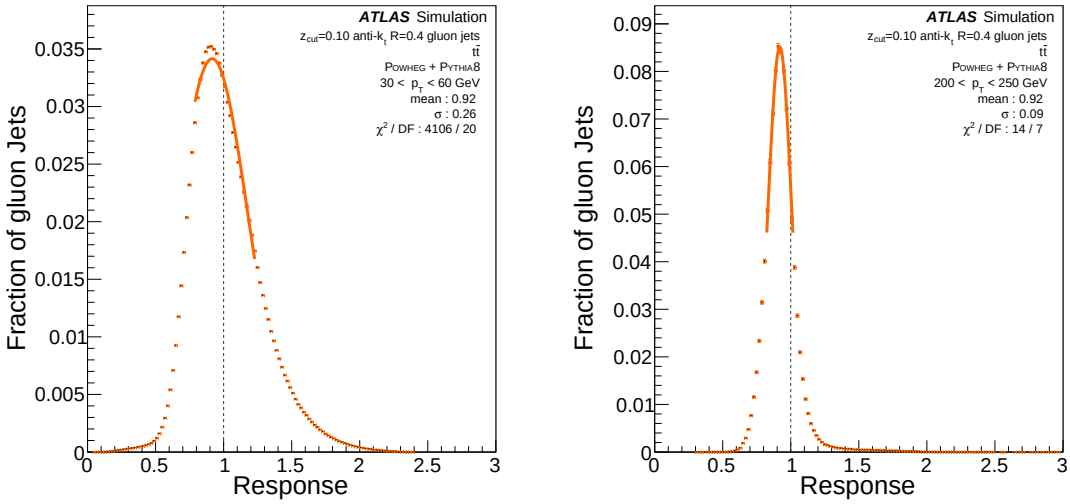
$$\frac{p_T{}^{grooming}_{reco-jet}}{p_T{}^{ungroomed}_{truth-particlejet}} \tag{6.3}$$

I bin these responses by $p_{\mathrm{T}}{}^{grooming}_{reco-jet}$, and fit a Gaussian function to the total result of that bin. I then find the mean value, standard deviation (σ) and χ^2 value of this fitted Gaussian function. The results of this process are presented for gluon-jets in Figure 6.74, and for all jet flavours in Figure F.1 and Figure F.2.

Looking at the response for jets with p_{T} between 30 GeV and 60 GeV, I can see that gluon- and c-jets have the mean response values closest to one, at a value of 0.99 and 1.01 respectively. b-jets are close to a value of one, with 0.98 and uds-jets are a little further at 1.06. I observe that tau-jets have means with the furthest values to one, with 3-pronged tau -jets and 1-pronged tau-jets at 1.08 and 1.15 respectively. For jets with p_{T} between 200 GeV and 250 GeV I also see that gluon-, b- and c-jets have values of mean closes to one, at 0.99, 1.01 and 1.01 respectively. I see uds-jets are again a little further away, at 1.02. I also see that tau-jets have means with the furthest value from one, with 3-pronged tau -jets and 1-pronged tau at 1.03 and 1.04 respectively. Over all I see that other than for gluon- and c-jets (with the closest mean in the first p_{T} bin), the mean is closer to one for those jets with p_{T} between 200 GeV and 250 GeV.

When looking at the standard deviation for jet response for jets with p_{T} between 30 GeV and 60 GeV, I see most flavours have values very close to one another, with uds-, 1-pronged tau - and 3-pronged tau -jets each with values of 0.2, b- and c-jets at values of 0.21, and gluon-jets somewhat broader at 0.24. I see a general trend of narrowing as I move to for jets with p_{T} between 200 GeV and 250 GeV. uds-, gluon-, b- and 3-pronged tau -jets all have a standard deviations of 0.07, and c- and 1-pronged tau-jets have a standard deviations of 0.08. All have become more narrow, particularly gluon-jets, an outlier in the first p_{T} bin.

Following the calculation of this for all the chosen p_T bins, I plot the means from each bin for comparison, the results for all five flavours are shown in Figure 6.75. Looking at the results for Mean jet response I see a general trend of mean values moving closer to one as the p_T of the jets increase – this is particularly true for those with the furthest values in the p_T bin 30 GeV to 60 GeV, uds-, 3-pronged tau - and 1-pronged tau-jets. I see that those jet flavours with the closes mean to one in both of the previous p_T bins, gluon- and c-jets, have the most stable mean throughout the range of p_T bins. I also see that in every case the mean for jets with a p_T between 0 GeV and 30 GeV, out lowest p_T bin, is significantly lower than the mean for jets within the next p_T bin, between 30 GeV and 60 GeV.



(a) p_T range 30 GeV to 60 GeV (b) p_T range 200 GeV to 250 GeV

Figure 6.76: Jet response of gluon-jets with soft drop grooming of strength Z $_{cut}$ = 0.1 applied, in the p_T range 30 GeV to 60 GeV (b) and 200 GeV to 250 GeV (b)

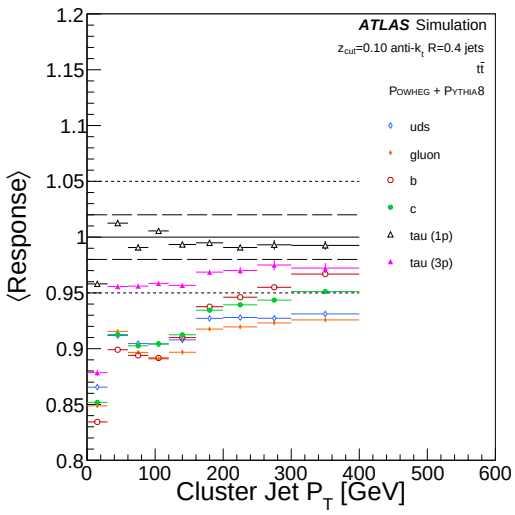


Figure 6.77: Mean jet response of all jet flavours with soft drop grooming of strength Z $_{cut}$ = 0.1 applied.

Following the application of soft drop grooming on the reco-jets, there are some interesting results. We show the results for gluon-jets in both bins in Figure 6.76, the results for "bad" bins of all flavours in Figure F.3 and for "good" bins of all flavours in Figure F.4. Looking first for the change in jet response for those jets of $30\,\text{GeV} \leq p_\text{T} < 60\,\text{GeV}$ first note that all the mean values are reduced. This is expected, as the reco-jet is losing p_T as clusters are groomed away, but the reference truth-particle jet is not. There is an average reduction in the mean of 11%. For tau-jets this brings the mean value closer to unity, especially in the case of 1-pronged tau-jets. For all non-tau flavours this leads to moving further away from unity, most by 7 or 8%. The standard deviation for all "bad" bins also increases by close to 15%.

Moving onto the "good" bins, with jets of $200\,\text{GeV} \leq p_\text{T} < 250\,\text{GeV}$, again the mean is reduced, with an average of 7%. Once again tau-jets move closer to unity, particularly 1-pronged tau, which had the highest value before. All other jet flavours move further away from unity by close to 6%. The effect that soft drop grooming has on tau-flavoured jets (and most obviously on 1-pronged tau-jets) indicates once again that there are significant contributions present in ungroomed reco-jets, that are not present in either reco-jets with soft drop grooming applied, or in truth-particle jets. The standard deviation for 1-pronged tau, uniquely, is reduced by 12.5% following application of soft drop grooming, with all other flavours increasing between 12 and 25%.

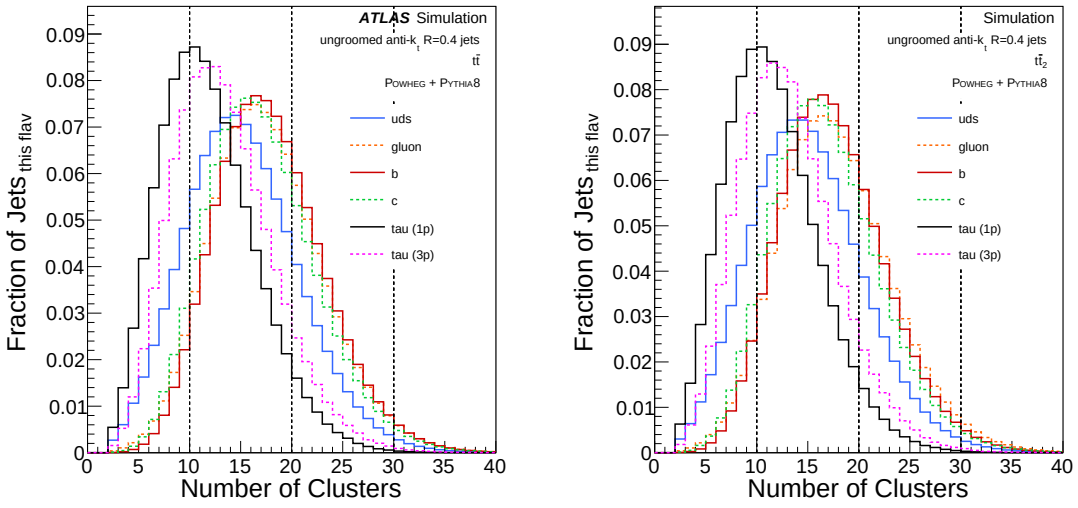
Ultimately it appears that the application of soft drop grooming to these reco-jets has not necessarily improved the resolution in reference to ungroomed truth-particle jets, other than in the case of tau-flavoured jets. The results for ungroomed reco tau-jets (and particularly 1-pronged tau()-jets) when compared to ungroomed truth-particle jets indicates the presence of extraneous energy that is removed definitely following application of soft drop grooming.

6.13 Soft Drop Grooming on b-tagged Jets in ATLAS Data

This set of studies will focus on analysis of the ATLAS data sample discussed in Section 6.2.1. Comparisons of JSS observables of flavour-tagged reco-jets undergoing soft drop grooming are undertaken made between this data sample and a number of MC simulation samples. The MC simulation samples that have been selected for these studies have weighted pile-up distributions to match conditions from ATLAS in 2018, as described in Section 6.2. The first of these studies explored the chosen MC simulation of $t\bar{t}$ events, known as "$t\bar{t}_2$", and how it compares to the $t\bar{t}$ samples used in the previous studies in this chapter.

6.13.1 Soft Drop Grooming with $t\bar{t}_2$

The first investigation undertaken for the MC simulation sample $t\bar{t}_2$ is a comparison between truth-flavour-tagged jets. In Figure 6.78 a comparison between the number of clusters in ungroomed truth-flavour-tagged jets can be seen for two samples, $t\bar{t}_2$ and $t\bar{t}$ (the MC simulation sample used through the previous studies in this chapter).



(a) Previous $t\bar{t}$ Sample

(b) New $t\bar{t}_2$ Sample

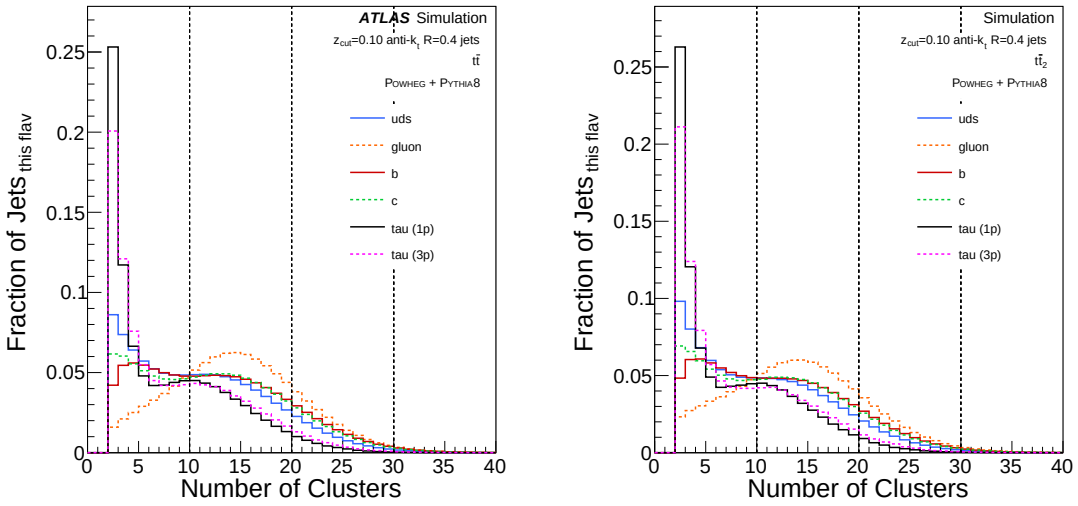
Figure 6.78: Number of clusters for ungroomed jets, from the previously used $t\bar{t}$ sample (a) and the new $t\bar{t}_2$ sample (b).

These two sets of distributions show the same pattern (with very minor differences) for number of clusters in each of the truth-flavour-tagged sets of reco-jets. Firstly, tau-jets, particularly 1-pronged tau-jets, have a narrow distribution with a high peak at the lowest number of clusters. Secondly, b-, c- and gluon-jets have broader distributions with peaks at higher number of clusters (and close to one another). Finally, uds-jets lie somewhere between the two.

Following the application of soft drop grooming with a strength of $Z_{cut} = 0.10$ to jets from each of these samples, it can be seen that this pair of distributions matches very closely, once again. This is shown in In Figure 6.79. In each case, tau-jets change the most following the application of grooming. Both sets of distributions show the tau-flavour-tagged jets, particularly 1-pronged, with very high peaks at two clusters, the minimum number of clusters allowed in a reco-jet. In each case, this suggests the reco-jets have a well-defined hard-core, such that wider-angled, softer clusters meet the requirement to be groomed away.

In both samples, gluon-jets appear to be almost unaffected compared to other flavours of jet. The peak of these distributions move from about 17 clusters to about 15, suggesting very few clusters are being groomed away. As discussed previously, this indicates that these gluon-jets lack a well-defined hard-core, and, therefore, there is no well-defined wide-angle, soft radiation to be groomed. Once again, the effect on quark-jets lies in-between; uds-jets are the most groomed of the three quark-flavours and b -jets are the least.

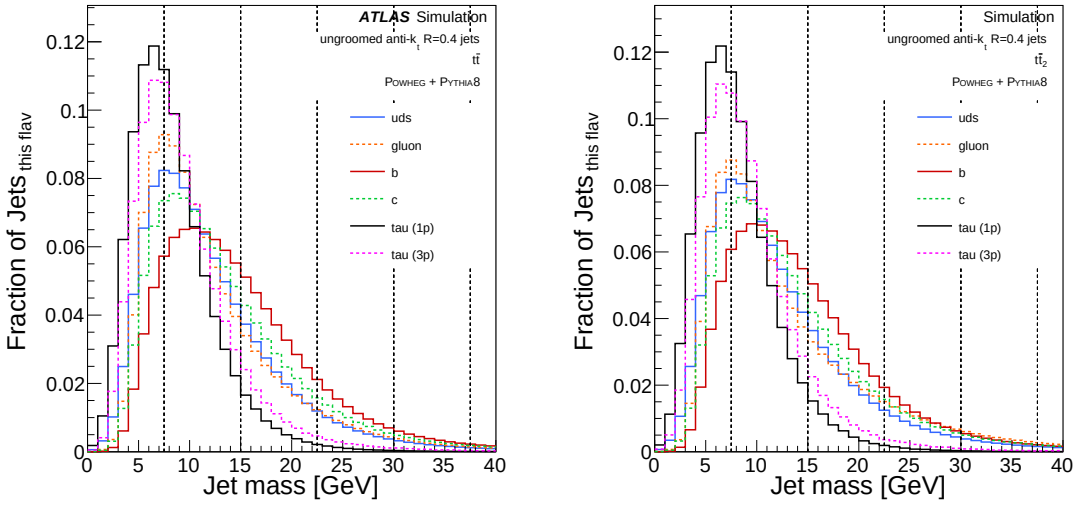
The same comparison can be made for jet mass of these jet flavours. Looking at the distribution of jet mass for these two sample in Figure 6.80, it can be seen that once again the results are very close. Jets that are flavour-tagged as tau-jets have a narrower distribution and have the lowest peak, at a value of 6 GeV. Through gluon-, uds- and c-jets onto b -jets, the distributions steadily get broader with lower peaks at higher values. The peak value of jet mass for b -jets in



(a) Previous $t\bar{t}$ Sample (b) New $t\bar{t}_2$ Sample

Figure 6.79: Number of clusters for jets soft drop groomed, with strength $Z_{cut} = 0.10$, from the previously used $t\bar{t}$ sample (a) and the new $t\bar{t}_2$ sample (b).

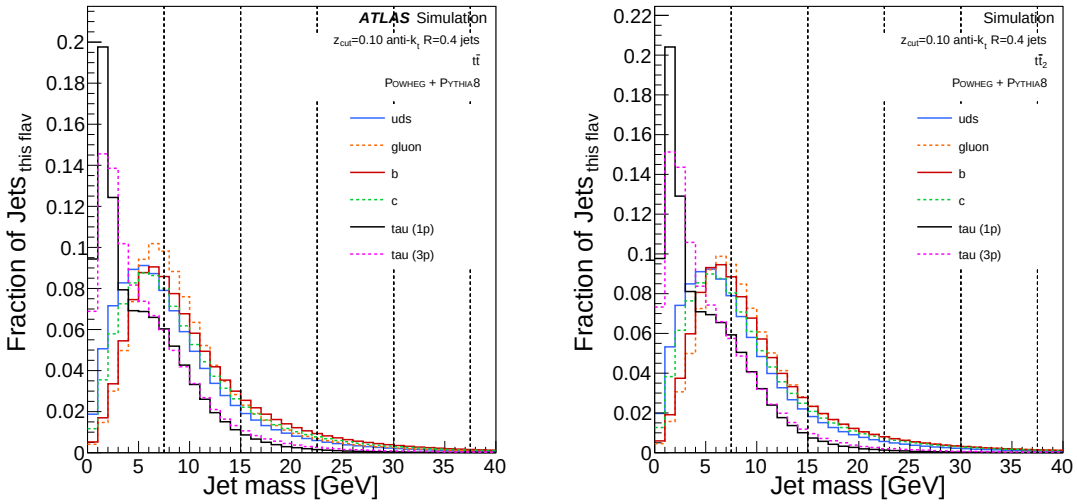
each sample is 9 GeV and 10 GeV for $t\bar{t}$ and $t\bar{t}_2$, respectively.



(a) Previous $t\bar{t}$ Sample (b) New $t\bar{t}_2$ Sample

Figure 6.80: Jet mass for ungroomed jets, from the previously used $t\bar{t}$ sample (a) and the new $t\bar{t}_2$ sample (b).

Applying soft drop grooming with a strength of $Z_{cut} = 0.10$ reveals a familiar pattern in Figure 6.81: the jet mass distributions for $t\bar{t}$ and $t\bar{t}_2$ are once again very similar. In both cases tau-jets have been most heavily groomed, now with a narrow peak at 1 GeV. This is followed by b-, c- and uds-jets, respectively, which now have distributions that are much more similar to one another, each now with a peak at 5 GeV or 6 GeV. The least groomed are gluon-jets, with a peak that has moved only from 7 GeV to 6 GeV.



(a) Previous $t\bar{t}$ Sample

(b) New $t\bar{t}_2$ Sample

Figure 6.81: Jet mass for jets soft drop groomed, with strength Z_{cut} = 0.10, from the previously used $t\bar{t}$ sample (a) and the new $t\bar{t}_2$ sample (b).

In each of these cases, the jet mass has not changed as much as the number of clusters by the application of grooming, indicating that those clusters groomed away were soft. The jet-flavours with the distributions most changed (i.e. the tau-flavoured jets) indicate that those flavour-tagged jets have more well-defined hard cores, such that softer clusters meet the requirement for grooming. Those jets with less change (such as the gluon-jets in particular), indicate a less well-defined hard-core.

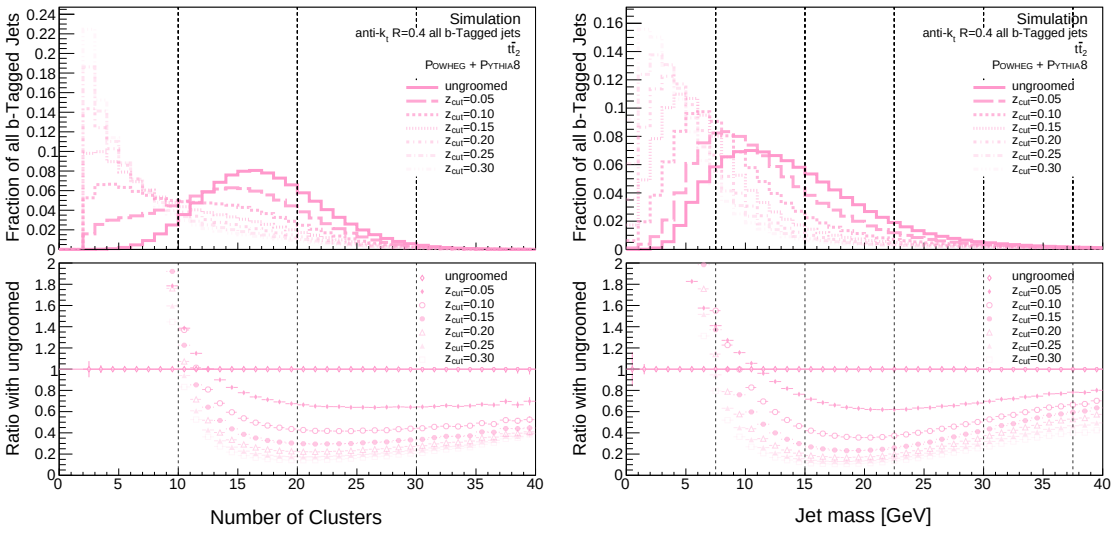
Ultimately, despite the fact that these two MC simulation samples are weighted to different pile-up distributions, $t\bar{t}$ matched to ATLAS data from 2017 and $t\bar{t}_2$ matched to data from 2018, the results they show are very similar. In all cases, this grooming reduces the number of clusters in jets, and decreases the jet mass to a lesser extent. The flavour-tagged jets from these two samples, both before and after the application of soft drop grooming, show the same patterns and behaviour as one another. In both samples, tau-jets show the most distinctive and dramatic effect of grooming, with many jets groomed down to the minimum of two clusters, indicating that the well-defined hard cores within these jets are distinct enough from soft, wide-angled radiation. This is particularly true for 1-pronged tau-jets. Gluon-jets in both of these samples are the least effected by grooming, suggesting much less well-defined internal structure. Those jets flavour-tagged as quarks, b, c and uds, lie between these flavours, with each more groomed that the last, respectively.

Soft Drop Grooming All b-Tagged Jets.

Truth-level information is not available in data from the ATLAS detector, only from MC simulation samples; therefore, for comparisons between the two to be made, a common jet-flavour-tagging method must be established. This will be achieved by using traditional b-tagging meth-

ods, as described in Section 6.4.3, as the only means of flavour-tagging any jets passing the standard selection process used so far.

Figure 6.82(a) displays the number of clusters for all b-tagged jets in the $t\bar{t}_2$ MC sample that pass the standard selection process after application of each strength of soft drop grooming. As can be seen, the general trend reflects that seen in jets of other flavours and from other samples. As the grooming strength is increased, the distribution for number of steadily narrows and shifts to a lower value. Ungroomed, the peak number of clusters for these jets is at 16, after a grooming strength of $Z_{cut} = 0.10$ has been applied, this peak has moved to 3 clusters and after a grooming of strength $Z_{cut} = 0.20$ the peak has already reached the minimum number of clusters required in a jet to be included, at two clusters. Comparing this to for truth-flavour tagged jets, it can see the results of grooming are most similar to those jets truth-flavour tagged as b -jets.



(a) Number of clusters

(b) Jet mass

Figure 6.82: The effects of different soft drop grooming strengths on number of clusters (a) and the jet mass (b) for all b-tagged jets in $t\bar{t}_2$.

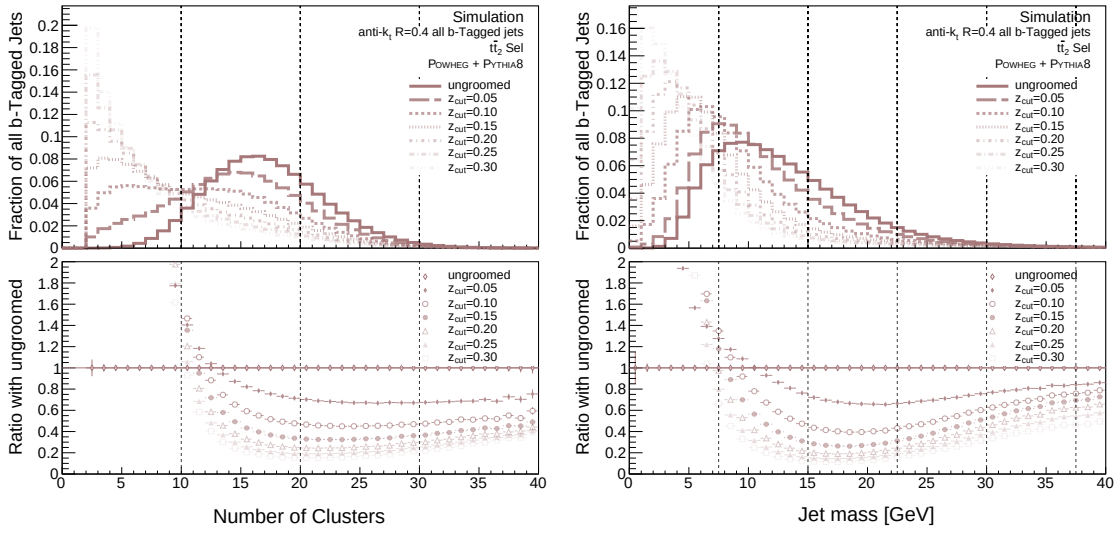
Figure 6.82(b) shows the jet mass of these same jets, once again, after different strengths of soft drop grooming has been applied. The distribution steadily narrows and shifts to a lower value as the grooming strength increases. This rate of change after applying grooming is less than that for number of clusters, indicating that those cluster that are groomed away are soft, and contribute proportionally little mass. Applying a soft drop grooming of $Z_{cut} = 0.10$ shifts the peak value of jet mass from 10 GeV to 6 GeV, once again this matches closely with the results for those jets truth-flavour tagged as b -jets.

6.13.2 Soft Drop Grooming with $t\bar{t}_2 Sel$

Following these exploratory studies with the sample $t\bar{t}_2$, an investigation the effects of soft drop grooming on JSS observables for all b-tagged jets in different samples was undertaken. One of

these samples is data from the ATLAS experiment, discussed in Section 6.2.1. In order to make a closer comparison between this data and the two MC simulation samples used in this section of the analysis, the event selection discussed in Section 6.3.2 was applied. This set of event selection criteria is used to select $t\bar{t}$ pair production events with leptonic and semi-leptonic decays. As no truth-level information is available in ATLAS data, the only flavour-tagging process I will use is traditional b-tagging methods, as described in Section 6.4.3. Any jets passing these selections and flavour-tagging will be studied here.

Firstly, I shall apply these event- and object-selections to the sample $t\bar{t}_2$. This new sample, with these selections applied, is labelled $t\bar{t}_2 Sel$. Figure 6.83 displays the number of clusters for all b-tagged jets in the $t\bar{t}_2 Sel$ MC sample following application of different soft drop grooming. Before the application of soft drop grooming, the distribution of number of clusters within all b-tagged jets looks to be similar to previous results. The distribution is fairly broad and peaks at 16 clusters. Following the application of soft drop grooming, the familiar change in distribution is seen: the peak becomes narrower and shifts to a lower number of clusters as they are groomed away from jets. After application of soft drop with a strength of $Z_{cut} = 0.10$, the peak is at 5 clusters. The peak reaches the minimum number of clusters necessary for selection, at two clusters, following a grooming strength of $Z_{cut} = 0.20$.



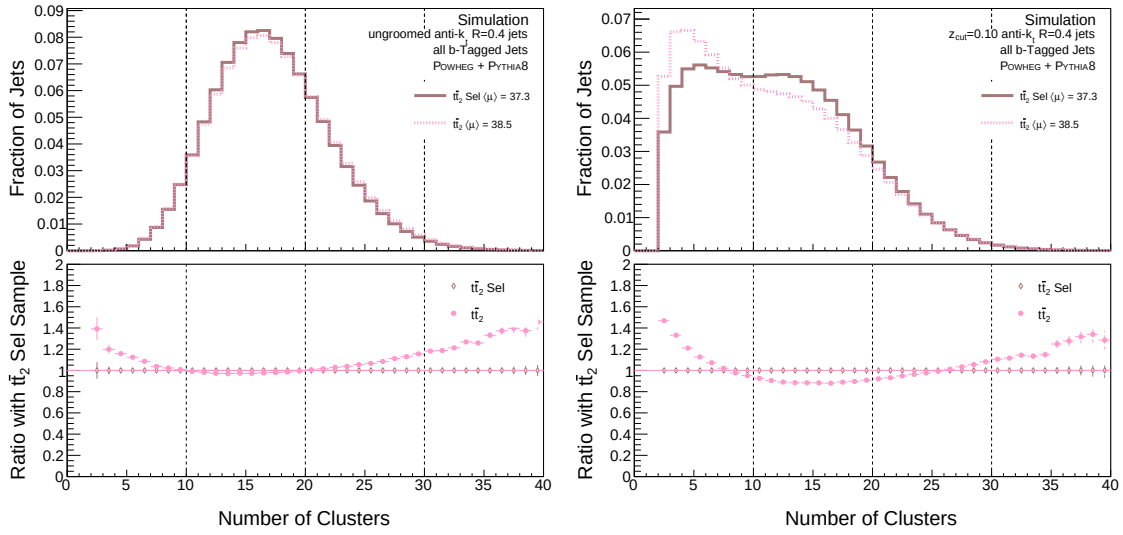
(a) Number of clusters (b) Jet mass

Figure 6.83: The effects of different soft drop grooming strengths on number of clusters (a) and the jet mass (b) for all b-tagged jets in $t\bar{t}_2 Sel$.

Figure 6.83 shows the effect on jet mass when applying soft drop grooming. As increasingly strong soft drop grooming is applied, the distribution of jet mass shifts to smaller values and becomes narrower. For each strength of soft drop grooming, however, this change is less than for number of clusters in these jets. This suggests that the clusters being groomed away have proportionally little mass compared to the clusters that remain within the jet. Initially jet mass has a peak at 9 GeV, but following application of soft drop grooming with a strength of $Z_{cut} = 0.10$, the peak value is at 6 GeV. Both sets of results, for number of clusters and jet mass,

is comparable to results for all b-tagged jets in $t\bar{t}_2$, as well as results for jets truth-flavour tagged as b -jets.

Figure 6.84 shows a direct comparison for the number of clusters in all b-tagged jets. This is shown both from $t\bar{t}_2$, where the standard selection processes used throughout this chapter has been applied, and from $t\bar{t}_2 Sel$, where the event- and object-selections have been chosen for direct comparison to ATLAS data. Before the application of soft drop grooming, these two distributions look very similar; however, applying soft drop grooming with a strength of $Z_{cut} = 0.10$ reveals some difference in behaviour between the two sets of jets. Jets from the sample with the additional selection criteria applied are less likely to have clusters groomed away. This indicates that they have a less well-defined hard-core. This new result reflects more closely the distribution of those jets that are truth-flavour tagged as b-jets.



(a) Ungroomed

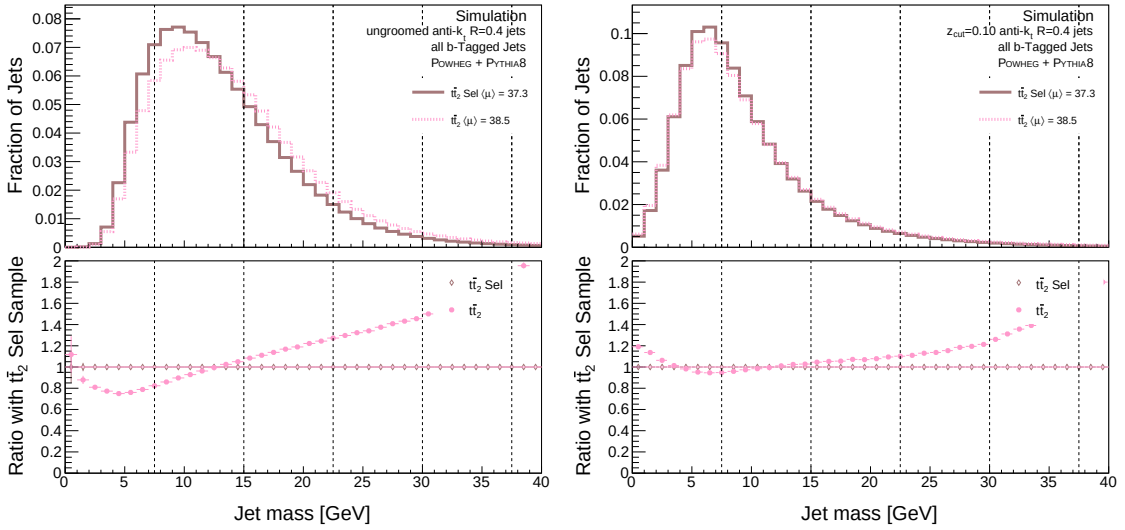
(b) Soft drop strength $Z_{cut} = 0.10$

Figure 6.84: Number of clusters for all b-tagged jets from both $t\bar{t}_2$ and $t\bar{t}_2 Sel$ samples with no soft drop grooming applied (a) and with soft drop grooming of strength $Z_{cut} = 0.10$ applied (b).

Displayed in Figure 6.85 is a comparison of jet mass for jets from these two samples. Initially, there is some difference between the two: the jet mass distribution for jets from $t\bar{t}_2 Sel$ is slightly narrower and peaks at a slightly lower value. Ungroomed jets from $t\bar{t}_2$ have a higher mass, despite having the same number of clusters, indicating that the clusters from jets in the $t\bar{t}_2 Sel$ sample are slightly softer. Following the application of soft drop grooming of strength $Z_{cut} = 0.10$, the distributions for jet mass look much more similar. Distributions for jet mass have changed less than for the respective distributions for number of clusters, reflecting the results seen in previous studes in this analysis.

6.13.3 Soft Drop Grooming with ATLAS Data and $t\bar{t}$ MC.

This section of the analysis will focus on all b-tagged jets from ATLAS data, discussed in Section 6.2.1. In order to make a closer comparison between this data and MC simulation samples



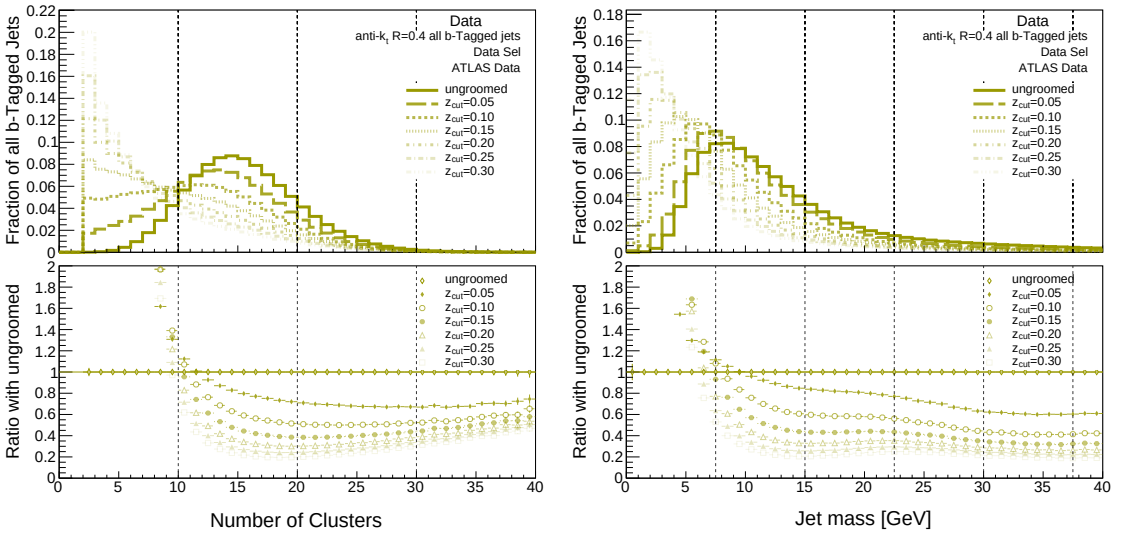
(a) Ungroomed (b) Soft drop strength $Z_{cut} = 0.10$

Figure 6.85: Jet mass for all b-tagged jets from both $t\bar{t}_2$ and $t\bar{t}_2 Sel$ samples with no soft drop grooming applied (a) and with soft drop grooming of strength $Z_{cut} = 0.10$ applied (b).

used in this section of the analysis, the event selection discussed in Section 6.3.2 is applied. This set of event selection criteria is used to select $t\bar{t}$ pair production events with leptonic and semi-leptonic decays [**?**].

The effects of soft drop grooming on number of clusters within all b-tagged jets from this ATLAS data sample can be seen in Figure 6.86(a). Each of the distributions seen lare comprable with previously observed results. Before grooming is applied, a broader distribution with a peak at 14 clusters is seen. As with jets from MC simulations, as successively stronger strengths of soft drop grooming is applied, the distribution becomes more narrow and shifts to a lower number of clusters. Following grooming of strength $Z_{cut} = 0.10$, the distribution has shifted to fewer clusters, and now peaks at a value of 10 or 11 clusters. Once a grooming strength of $Z_{cut} = 0.15$ is reached, the distribution starts to peak at the minimum number of clusters for a jet, at two clusters. When comparing to results from the $t\bar{t}$ MC simulation samples previously studied in this analysis, this distribution appears to lie somewhere between those for jets truth-flavour tagged as b -jets and those truth-flavour tagged as gluon-jets.

Figure 6.86(b) shows the effect of soft drop grooming on the mass of all b-tagged jets from ATLAS. As before, the general trend of grooming effects can be seen here the same as in previous results: successively higher strengths of soft drop grooming makes the distribution of jet mass more narrow and shifts it to lower values. This change is slower and more steady than the change in number of clusters, suggesting that the cluster groomed away are making proportionally-small contributions to the mass of the jet. Before any soft drop grooming is applied, the distribution peaks at 7 GeV. This peak moves to a value of 5 GeV following application of soft drop grooming strength $Z_{cut} = 0.10$. These distributions and changes from soft drop grooming seem to lie somewhere between those for truth-flavour tagged gluon-jets and the quark-jets (uds, c, b) from the MC simulated studies previously.



(a) Number of clusters

(b) Jet mass

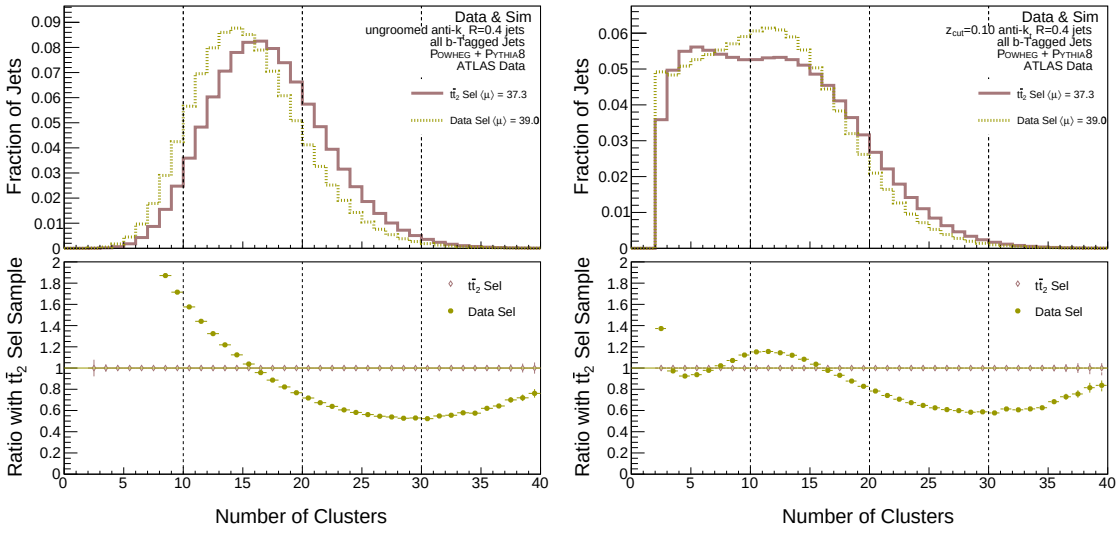
Figure 6.86: The effects of different soft drop grooming strengths on number of clusters (a) and the jet mass (b) for all b-tagged jets in ATLAS data.

A direct comparison between all b-tagged jets from ATLAS data can be made to jets from the $t\bar{t}$ MC simulation sample with the same event- and object- selection applied, $t\bar{t}_2Sel$. First, the distributions for number of clusters, shown in Figure 6.87 is compared. Before soft drop grooming is applied, the shape of the two distributions are very similar, the only notable difference is that all b-tagged jets from $t\bar{t}_2Sel$ have a slightly higher number of clusters than those jets from data. The distribution for data peaks at a value of 14 clusters, whereas for $t\bar{t}_2Sel$ it is 16. This suggests that the jets that compose our $t\bar{t}_2Sel$ sample are more likely to possess a higher number of clusters.

Following the application of soft drop grooming with a strength of Z $_{\text{cut}}$ = 0.10, it can be seen that jets from $t\bar{t}_2Sel$ are more likely to be groomed, suggesting they have a more well-defined hard-core. The distribution for number of clusters from $t\bar{t}_2Sel$ now peaks at a value of 5 clusters. This is closer to the results for truth-flavour tagged b -jets than those jets from data are. Jets from data now peak at a value of 10 clusters. They seem to lie somewhere between jets truth-flavour tagged as b -jets and gluon-jets.

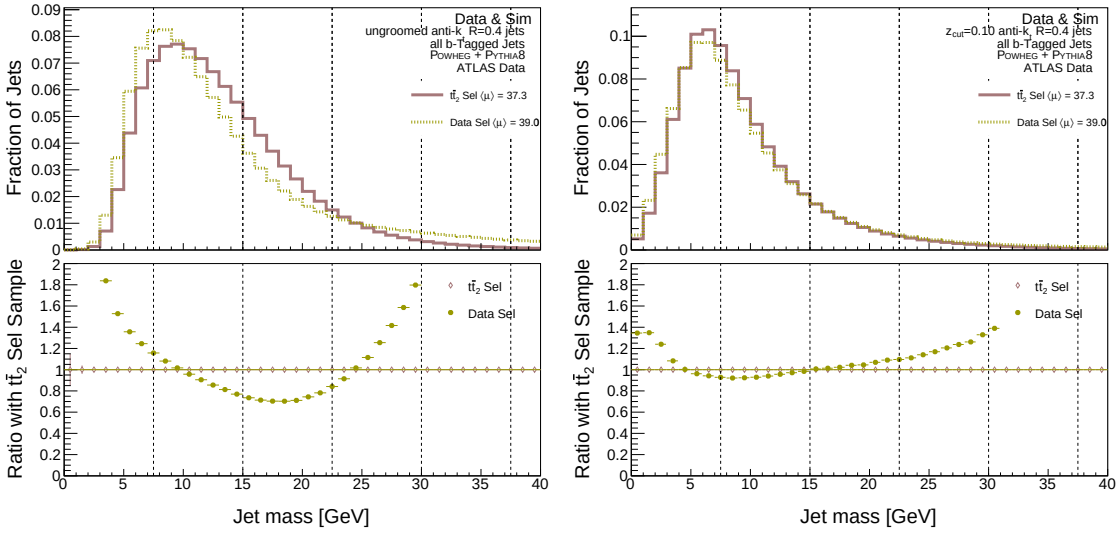
This same comparison can be made for values of jet mass, shown in Figure 6.87. Before grooming is applied, the two distributions are very similar. Jets from data peak at a slightly lower value of 7 GeV, which is closes to the value for jets from MC samples truth-flavour tagged as uds- or gluon-jets, and the distribution for jets from $t\bar{t}_2Sel$ peaks at 9 clusters, which is the same as jets from MC samples truth-flavour tagged as b -jets. The similarities and slight difference between ungroomed jets from these two samples seem to suggest a similar cluster to mass ratio to one another.

After a soft drop grooming of strength Z $_{\text{cut}}$ = 0.10 has been applied, proportionally less mass has been removed. This indicates that those clusters removed were proportionally soft.



(a) Ungroomed (b) Soft drop strength $Z_{cut} = 0.10$

Figure 6.87: Number of clusters for all b-tagged jets from both $t\bar{t}_2$ and ATLAS data samples with no soft drop grooming applied (a) and with soft drop grooming of strength $Z_{cut} = 0.10$ applied (b).



(a) Ungroomed (b) Soft drop strength $Z_{cut} = 0.10$

Figure 6.88: Jet mass for all b-tagged jets from both $t\bar{t}_2$ and ATLAS data samples with no soft drop grooming applied (a) and with soft drop grooming of strength $Z_{cut} = 0.10$ applied (b).

This is especially true for jets from $t\bar{t}_2Sel$, which lost noticeably more clusters, but only slightly more mass. This is indicative of $t\bar{t}_2Sel$ having a more well-defined hard-core. Following the grooming, both distributions look more similar to one another, and are now very close. The distribution for jets from data now peaks at a value of 5 GeV, and is comparable to jets from the MC samples that are truth-flavour tagged as uds- and c-jets. The distribution for jets from $t\bar{t}_2Sel$, however, peaks at a value of 6 GeV, which matches those jets truth-flavour tagged as b- or gluon-jets. Overall these results seem to indicate that b-tagged jets from $t\bar{t}_2Sel$ and ATLAS data are similar to one another, with some differences. Jets from $t\bar{t}_2Sel$ seem to have a slightly more well-defined hard-core than jets from data and are very similar to those jets that are truth-flavour tagged as b -jets. This would suggest that b-tagged jets from this $t\bar{t}_2Sel$ sample are dominated by jets that would be truth-flavour tagged as b -jets, as would be expected from a combination of $t\bar{t}$ pair production (where 90.25 % of decays produce 2 b -jets and 9.5 % of decays produce 1) with b-tagging applied (where 60 % of b -jets are found and other flavours are rejected).

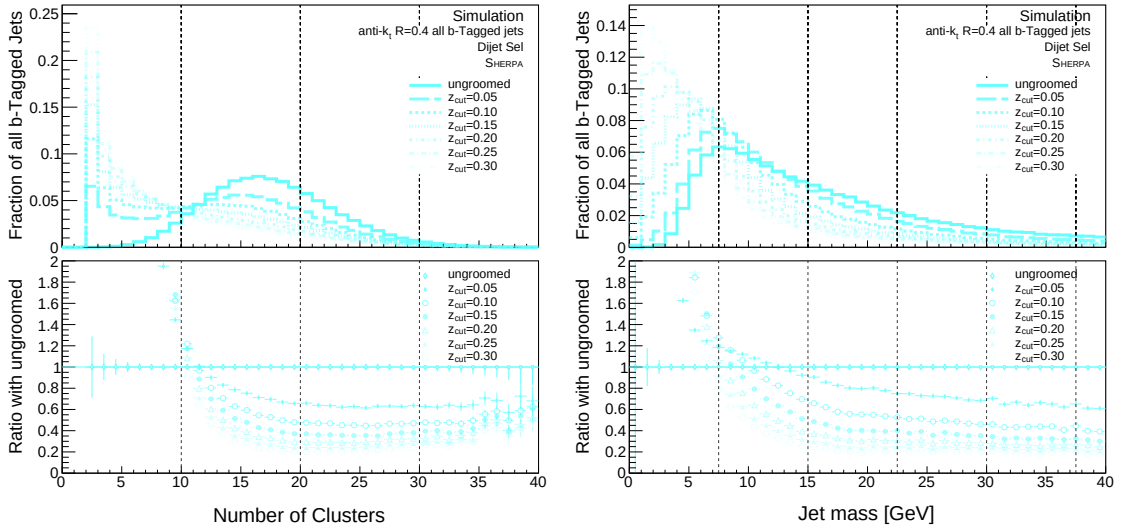
Those jets from data that are b-tagged seem to behave somewhere between jets from MC simulation samples that are truth-flavour tagged as b -jets and gluon-jets. This suggests that, despite the application of both a $t\bar{t}$ event selection and b-tagging object-selection, the composition of these remaining b-tagged jets from ATLAS data involves a significant portion of non-b-jets that are most likely gluon-jets. The results also indicate small, but noticeable contributions from other flavours such as uds- or c-jets.

6.13.4 Soft Drop Grooming with ATLAS Data and $t\bar{t}$ and Dijet MC.

In order to more thoroughly assess the results attained so far from ATLAS data, a final comparison was made to the MC simulation sample of dijet production events, discussed in Section 6.2.2. First the effects of applying soft drop grooming to all b-tagged jets from this MC sample was explored. This dijet sample has been selected with the same event- and object-selections as $t\bar{t}_2Sel$ and the ATLAS data (this selection is discussed in Section 6.3.2.

Figure 6.89(a) shows the effect of different strengths of soft drop grooming on the number of clusters in all b-tagged jets from this dijet sample. As successively stronger soft drop grooming is applied, the distribution number of clusters narrows and the peak value shifts quickly to the minimum value of two clusters. Before grooming, this peak is at 16 clusters, but after the application of $Z_{cut} = 0.05$, the "weakest" level of grooming applied, the peak has already moved to 2, with a second, slightly lower peak at 10 cluster. Once application of strength $Z_{cut} = 0.10$ has been applied, the first peak at two clusters is by far the more dominant. This indicate that after grooming, those jets that remain have a well-defined hard-core with those proportionally softer clusters that were initially included being groomed away. This result looks like a mix of the different truth-flavour tagged jets studied previously in $t\bar{t}$ samples, particularly with contributions from tau- or uds- jets, which are most groomed at this strength.

Figure 6.89(b) shows the effect of soft drop grooming on the mass of these jets. The effect



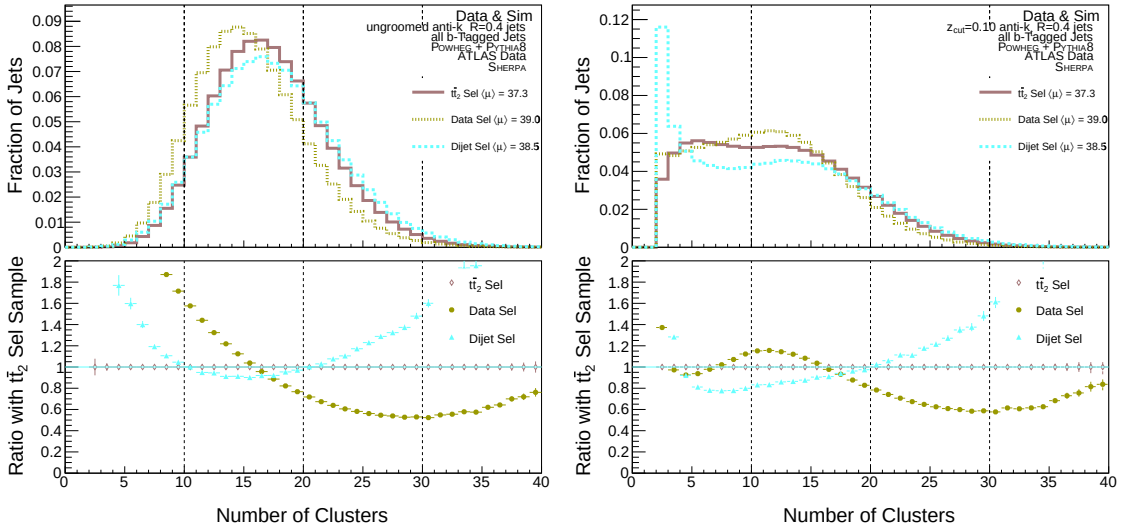
(a) Number of clusters

(b) Jet mass

Figure 6.89: The effects of different soft drop grooming strengths on number of clusters (a) and the jet mass (b) for all b-tagged jets in the dijet MC simulation sample.

of soft drop grooming on jet mass is that the distributions become narrower and shift to lower values of jet mass. This effect is less than that for number of clusters. This indicates, as seen before, that the clusters removed by this grooming are soft, and are contributing proportionally less to the mass of the jets. Before grooming is applied, the peak of the distribution is at 7 GeV. Following the application of soft drop grooming strength $Z_{cut} = 0.10$, this peak has moved to 6 GeV. As before, these results for b-tagged jets in the dijet sample, with a $t\bar{t}$ selection applied, look like a mix of the different distributions for truth-flavour tagged jets from $t\bar{t}$ MC samples

Next direct comparison of these results with all b-tagged jets from ATLAS data and from the $t\bar{t}_2 Sel$ MC sample was done. The first of these comparisons is shown in Figure 6.90(a). Before grooming is applied, the distributions for number of clusters are all similar in shape. The two MC samples peak at slightly higher values, perhaps indicating better cluster finding. The dijet peaks at the highest value, indicating that b-tagged jets from this sample contain more clusters. The different effects of applying soft drop grooming of strength $Z_{cut} = 0.10$ is shown in Figure 6.90(b). Applying this grooming effects the b-tagged jets in the dijet sample much more than from the other two samples, indicating more jets with well-defined hard cores or more soft clusters at a wide angle. Number of clusters for b-tagged jets from data are the lest effected of the three sets, indicating jets here have less well-defined hard-cores. Jets from $t\bar{t}_2 Sel$ behave very similarly to those jets truth-flavour tagged as b -jets, and therefore the composition of these b-tagged jets from $t\bar{t}_2 Sel$ are likely to be mostly correctly tagged b -jets. The difference in the other samples indicate that their compositions are not as "pure" and contain more jets that have been "mistagged" as b -jets. Jets from data appear to behave similar to a composition of b- and light- flavoured jets, such as gluon or uds. Jets from the dijet MC sample appear to show attributes similar to a blend of all flavours, with some jets being very heavily groomed (similar to tau- and uds-jets) and some less so (more similar to gluons).



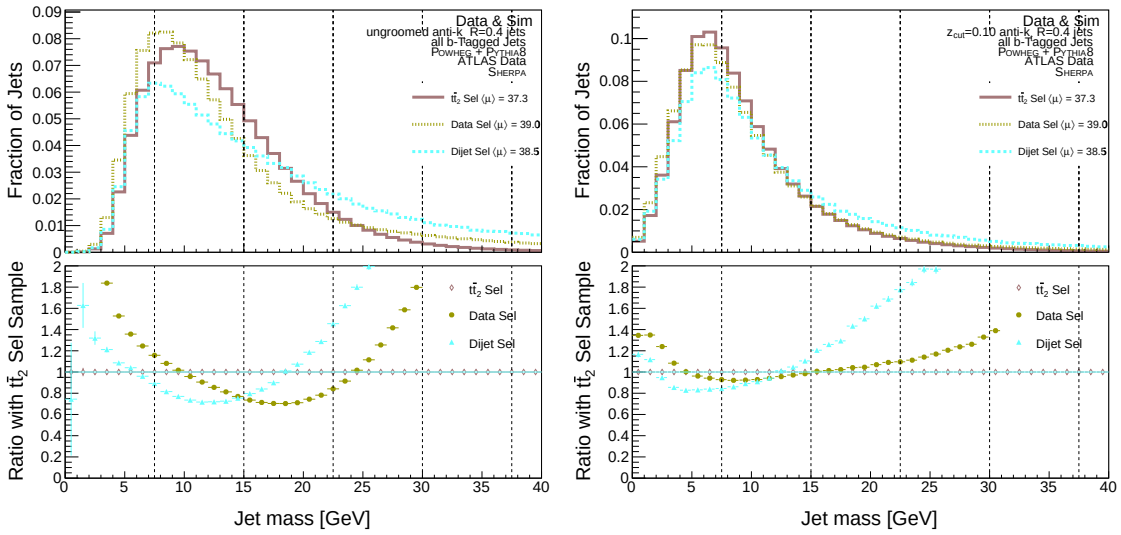
(a) Ungroomed (b) Soft drop strength $Z_{cut} = 0.10$

Figure 6.90: Number of clusters for all b-tagged jets from both $t\bar{t}_2$ and dijets MC simulation samples and ATLAS data samples with no soft drop grooming applied (a) and with soft drop grooming of strength $Z_{cut} = 0.10$ applied (b).

The same comparison can be made for jet mass, see Figure 6.91(a). Before grooming is applied, the distributions are similar in shape, but with some key differences. Jets from the $t\bar{t}_2 Sel$ sample are likely to have a higher mass that those jets from dijet or data. The distribution of $t\bar{t}_2 Sel$ is, once again, very similar to that of truth-flavour tagged b -jets from previous studies, suggesting that many of these are b-tagged jets. The distribution for dijet is most different from $t\bar{t}_2 Sel$, and data, with a noticeably broader distribution and with its peak at the lowest value. Once these jets have been groomed, the distributions for jet mass look closer, see Figure 6.91(b). In each case, the change in jet mass is less than the change in number of clusters, indicating that those clusters that have been groomed away were soft, and contributed proportionally less mass to jets. Here it is harder to harder to distinguish between possible contributions from underlying truth-flavour tagged jets, as the distributions from b-, c-, uds- and gluon-jets each look similar, although a large contribution from tau-jets can be discounted, as their distribution is obviously different.

6.13.5 Conclusions from Soft Drop Grooming with ATLAS Data.

In summary, these results display the the distributions for number of clusters and jet mass for all b-tagged jets taken from $t\bar{t}$ and dijet MC samples and from ATLAS data (each with a matching $t\bar{t}$ event- and object-selection applied). It was observed that these jets will behave somewhat similarly under application of soft drop grooming, but with some key differences. Distributions for jets from $t\bar{t}_2 Sel$ look a lot like distributions for jets truth-flavour tagged as b -jets. This would suggest that b-tagged jets from this $t\bar{t}_2 Sel$ sample are dominated by jets that would be truth-flavour tagged as b -jets, as would be expected from a combination of $t\bar{t}$ pair production (where 90.25 % of decays produce 2 b -jets and 9.5 % of decays produce 1) withan additional $t\bar{t}$



(a) Ungroomed (b) Soft drop strength $Z_{cut} = 0.10$

Figure 6.91: Jet mass for all b-tagged jets from both $t\bar{t}_2$ and dijets MC simulation samples and ATLAS data samples with no soft drop grooming applied (a) and with soft drop grooming of strength $Z_{cut} = 0.10$ applied (b).

selection and b-tagging applied (where 60 % of b -jets are found and other flavours are rejected).

Results from ATLAS data appear to lie somewhere between those for jets truth-flavour tagged as b- and gluon-jets. This indicates that applying $t\bar{t}$ selection and b-tagging has found a b -jet rich set of jets; however, the frequency ,and subsequent mistagging, of other jet flavours, such as gluon or light jets, is still common enough to effect the results.

Jets from the MC sample seem to behave in a way that indicates a mix of flavours. The application of $t\bar{t}$ event-selection and b-tagging is still likely to be rejecting those types that are would not be truth-flavour tagged as b -jets, but the frequency of these other jet types must be high enough to still contribute in a way that visibly shapes results. This is as expected, as dijet events will be dominated by jets from light quarks and gluons.

Overall, it appears that the use of soft drop grooming on b-tagged jets from different samples leads to some difference in results. Therefore, studying JSS observables following the application of soft drop grooming could be a useful tool in identifying the source of jets especially when combined with relevant event- and object-selections, such as use of traditional b-tagging methods.

CONCLUSIONS

Nothing has such power to broaden the mind as the ability to investigate systematically and truly all that comes under thy observation in life.

Marcus Aurelius

The observation of the Higgs boson in 2012 opened the gateway to exploring and testing new physics at the LHC. A decisive and rigorous test of the predictions of the SM will be the investigation of the Yukawa coupling between its heaviest particle, the top quark, and the Higgs Boson. A direct measurement of this Yukawa coupling is possible in the $t\bar{t}H$ production channel. Background at the LHC is dominated by multijet production, meaning the understanding of jet physics is critical to analyses in ATLAS. The importance of jet understanding is underlined by the nature of the Higgs decay channel with the highest branching ratio, ($H \to b\bar{b}$), and its dominant background, $t\bar{t}$ pair production with additional hadronic jets. All of these processes involve significant hadronic output, and of particular importance is an understanding of b-jets.

Optimisation of jet reconstruction, both in the ATLAS trigger and in offline analyses, relies on development of innovative techniques and processes, such as improved jet-tracking in the trigger, and utilisation of JSS and novel grooming methods offline.

Studies of potential trigger developments, such as implementation of the FTK, are invaluable in illuminating pathways to potential future improvements. This is increasingly true as accurate jet reconstruction becomes more complex due to higher levels of pile-up at the LHC.

The need to address shortcomings in traditional jet grooming algorithms, such as the handling of NGL, gives novel methods like soft drop grooming a unique advantage over the alternatives.

To investigate the potential benefits of soft drop grooming, analyses were undertaken with different strengths applied to flavour-tagged small jets, reconstructed from either calorimeter clusters or truth particles. This process was applied to jets both from $t\bar{t}$, $t\bar{t}H$ and dijet MC simulation and from ATLAS data. Different variables from these jets, including the JSS observables number of clusters and jet mass, were analysed before and after application of soft drop

grooming.

Both the jet mass and number of calorimeter clusters in these jets are reduced following this grooming. Common patterns appear for each of the different jet flavours and samples analysed, but with some key distinctions. Proportionally less jet mass is removed by soft drop grooming than the number of clusters that are groomed away. This is due to soft drop grooming removing soft, wide-angled radiation, with lower mass contributions, from these jets. Jets with differing truth-flavour tags are affected to different extents by the same grooming strength, revealing differences in the internal structure of the jets.

Application of soft drop grooming has a particularly strong effect on those jets flavour-tagged as tau-jets, reflecting the prominence of prong-like structure within jets, and their contrast with softer components. One effect of this is the increased likelihood in tau-jets to be groomed below two clusters – particularly 1-pronged tau . Gluons demonstrate the opposite behaviour, only presenting changes from soft drop grooming at higher strengths. This is indicative the less well-defined hard structure within the jet. Applying different strengths of soft drop grooming reveals thresholds for the various jet flavours to be more or less affected by the grooming.

These varying reactions to grooming can by employed to differentiate between jets of different origin. The first area in which this has been preliminarily successful is the distinction between "real" and "fake" tau jets. The distinctive features shown by tau-jets also looks to be a promising tool for improving understanding of tau-jet-finding at ATLAS. The accentuated tau-jet features following grooming can enhance differences from other jet flavours, and may assist in finding "lost" tau-jets, missed by traditional tagging methods. The study also looks particularly at jets from gluons and jets from u, d and s quarks. Separation of these different "light" jet flavours in these MC samples can also be improved following grooming.

An investigation into the effects of soft drop grooming on those jets truth-flavour tagged as b-jets that have either been found or missed by traditional b-tagging methods was also undertaken. These studies may inform how the use of soft-drop grooming could be used to understand the differences in these jets. This also enabled exploration of b-jets from $t\bar{t}$ and $t\bar{t}H$ MC simulations. Little difference is seen in the results here, as expected, but improving understanding of how these jets behave under grooming will inform future studies.

Further to this, comparisons of jets constructed from truth-particles with those constructed from calorimeter clusters are made. Applying soft drop grooming of strength $Z_{cut} = 0.05$ to cluster jets generally improves agreement with truth-particle jets, most evident in quark-jets. In gluon-jets, however, a strong agreement before application of soft drop grooming is observed, and they remain best ungroomed. Also demonstrated is the effects of soft drop grooming on the truth-particle jets themselves. It is seen that they are less altered by the grooming than cluster jets, indicating less contamination from soft-radiation. The effects of soft drop grooming on JES and JER, is explored. What is observed is that, generally, these tend to worsen following grooming. This is likely due to the differing effect on reconstructed jets and truth jets.

The final analysis section makes comparisons between 2018 data from the ATLAS experiment and two different MC samples: $t\bar{t}$ and dijet. An additional $t\bar{t}$ selection is applied before this comparison, and all reconstructed jets b-tagged by traditional methods are studied. Familiar patterns are seen in the distributions of JSS observables, number of clusters and jet mass. Differences in how jets from these samples behave indicate the varying composition of those jets b-tagged from each. Jets from the $t\bar{t}$ sample, $t\bar{t}_2 Sel$, behave very similarly to jets that have been truth-flavour tagged as b-jets. This is indicative of the dominance of b-jets in this sample. Distribution from both ATLAS data and the dijet MC sample suggest a more complex composition, consisting of varying flavours of jets, some of which having been mistagged as b-jets.

Statistical uncertainty of any result gained could ultimately be improved by increasing statistics – more MC sample events and more data from the ATLAS detector would lead to greater reliability in results for both trigger and soft drop studies.

Future studies of soft drop grooming with small jets could also include more JSS variables, potentially combined with multivariate analysis with and use of machine learning techniques such as a BDT. This could also improve comparison of jets found through traditional tagging methods (b-tagging or identification of τ leptons) to truth-flavour tagged jets in MC samples. Studies on jet response could be improved by including jet calibration stages in jet reconstruction.

Ultimately, more time and further studies following these investigation will hopefully lead to a greater understanding of the possibilities of utilising using soft drop grooming and JSS variables to improve understanding of jet reconstruction, searches for the Higgs and, therefore, the search for physics beyond the SM.

During this thesis I have covered two analyses. I have shown that the tighter clustering about the matched-jet centre, lead track and primary vertex in the jets built from HLT tracks is not present in jets built from FTK tracks, and propose this is due to the global nature of FTK track-finding. This result is much closer to results seen from the more detailed process of Offline track-finding in events, and leads to benefits, such as more low-p_{T} tracks being found in FTK events than HLT events. These differences are likely to improve tracking within jets, ultimately improving the efficiency and reliability of jet reconstruction in ATLAS. In particular, the ability to rapidly identify b-jets has the potential to allow lower p_{T} thresholds for jet triggers; this will be very helpful for analyses such as $H \to b\bar{b}$ [202]. Improved track finding is also crucial for countering the effects of pile-up, which will be a major challenge in the future high luminosity running of the LHC.

I have also shown the results of my studies using the soft drop grooming method on reconstructed jets to give further insight into the properties of jets from different origins. My studies of tau-jets showed for the first time that the soft drop groomed jet mass can assist in improving the purity of tau jet samples. Utilising the soft-drop jet mass alongside existing tau jet identification methods may help identification of more tau jets, whilst rejecting higher levels of background that come with using a looser working point.

The ability to discriminate between jets initiated by light quarks and those initiated by gluons is a potentially powerful tool in searches for new physics. I have shown that the soft drop jet mass has some discrimination power between light quark and gluon initiated jets.

Correctly identifying b-jets at ATLAS is of great importance, particularly in terms of Higgs and top quark measurements. Current b-tagging methods, when implemented at a high efficiency working point, suffer from impurity mainly from misidentified c-jets. I have shown that the soft drop jet mass is an effective new tool in distinguishing between c and b jets, with potential for inclusion in the existing multivariate ATLAS flavour tagging methods.

I have also shown that the soft drop jet mass for b-jets is consistent for several configurations across the different complex final states of $t\bar{t}$ and $t\bar{t}H$, as predicted by the POWHEGMC generator.

Finally, I have compared the number of clusters in, and the jet mass of, b-tagged jets from ATLAS data and MC simulation samples following of application of $t\bar{t}$ selection criteria and soft drop grooming. I observed a number of key similarities and differences between these samples, and demonstrated that soft drop grooming has potential to become a useful tool in the search for b-jets at ATLAS.

Ultimately, this all leads to a greater understanding of jet physics in ATLAS and can aid future searches for new physics beyond the SM.

A Soft drop grooming on all flavours of flavour-tagged small jets.

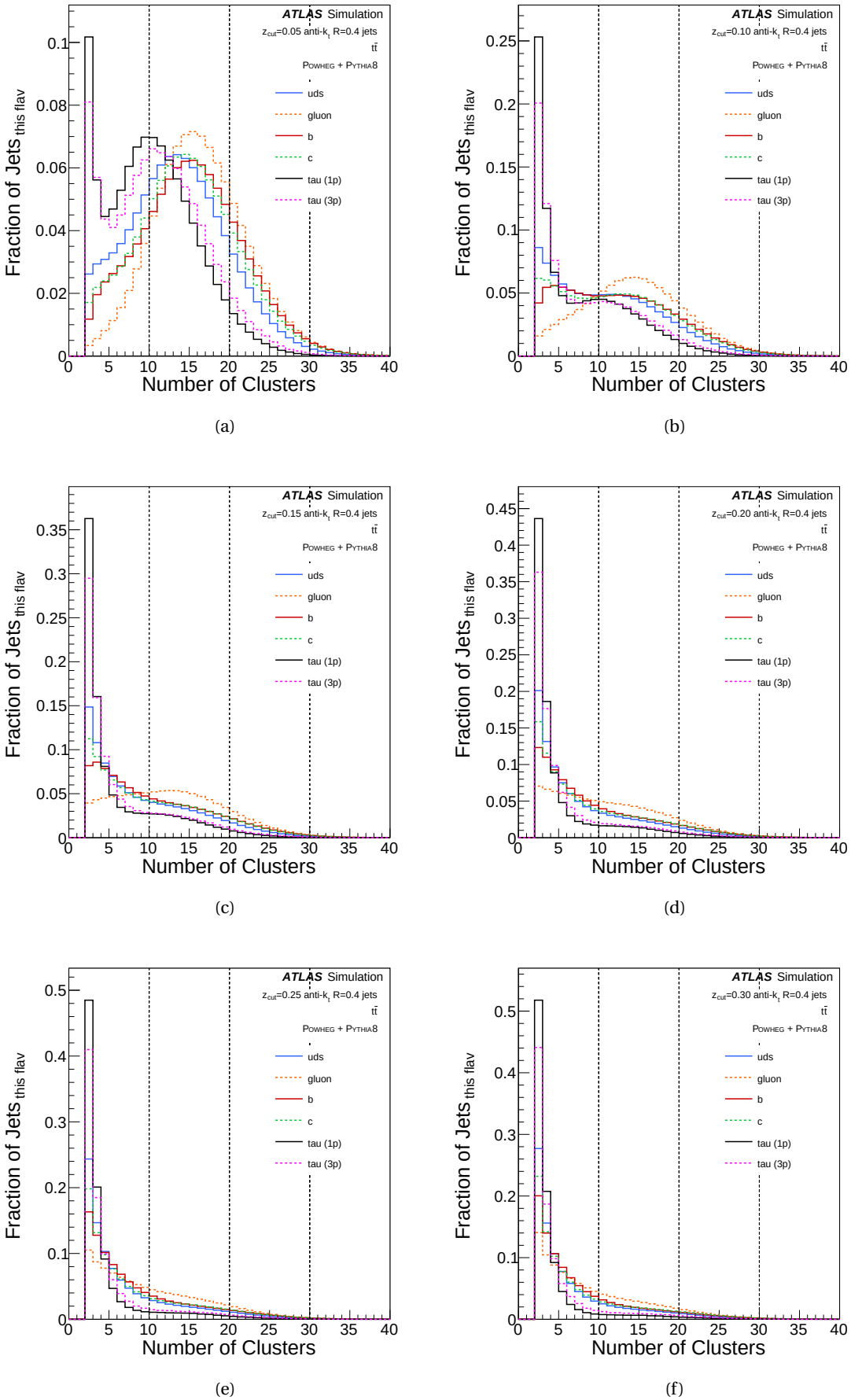


Figure A.1: Number of Clusters for all flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

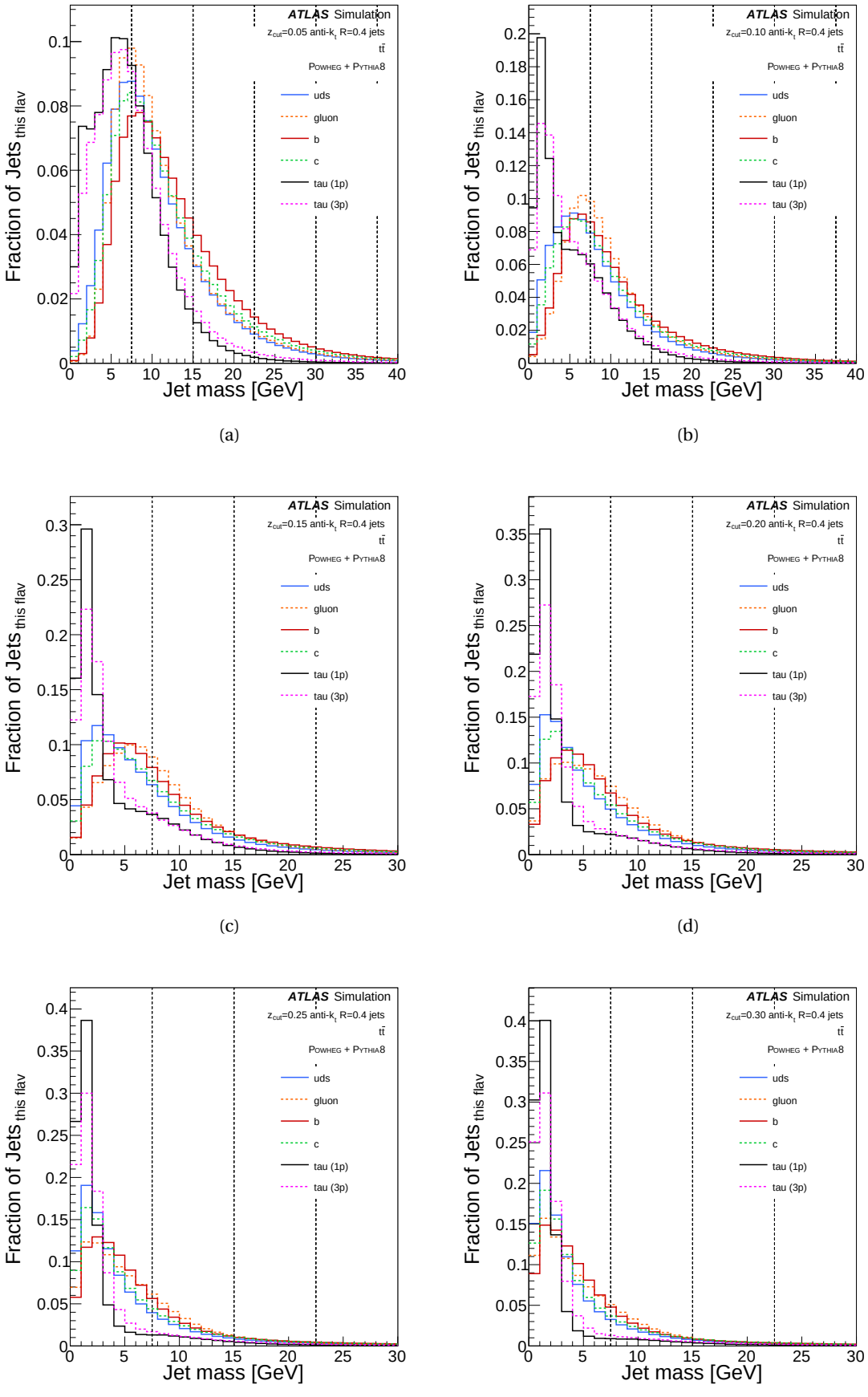
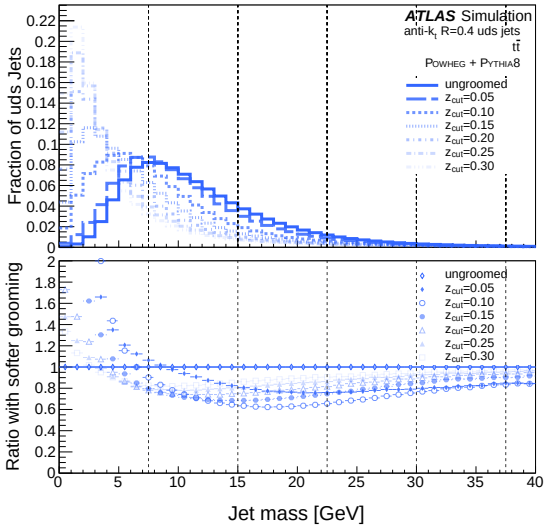
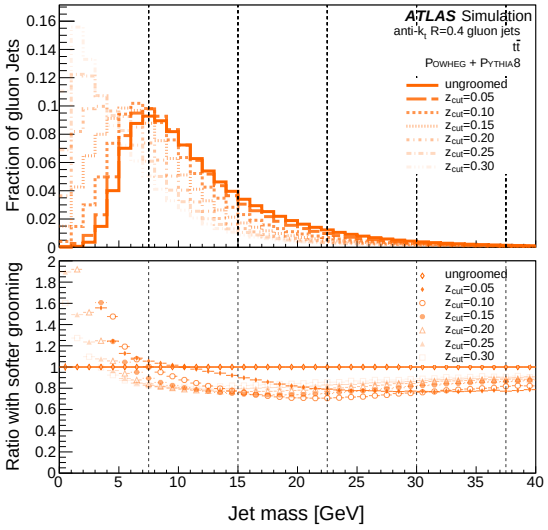


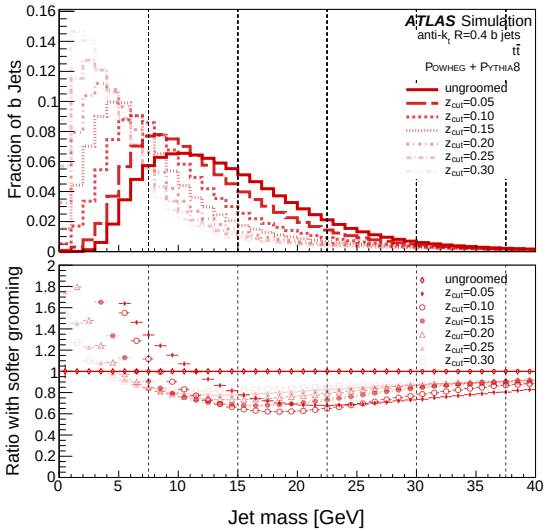
Figure A.2: Jet Mass for all flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)



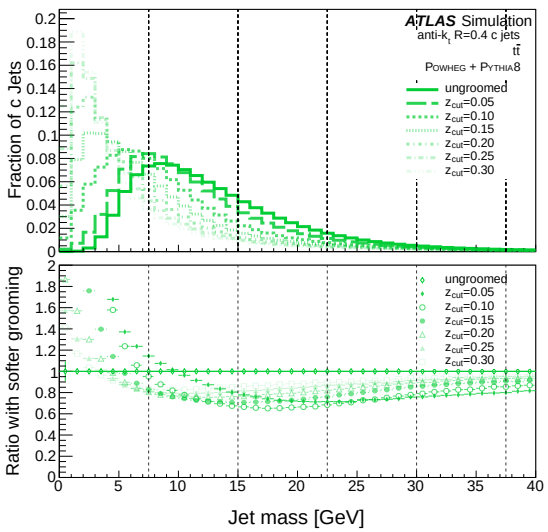
(a)



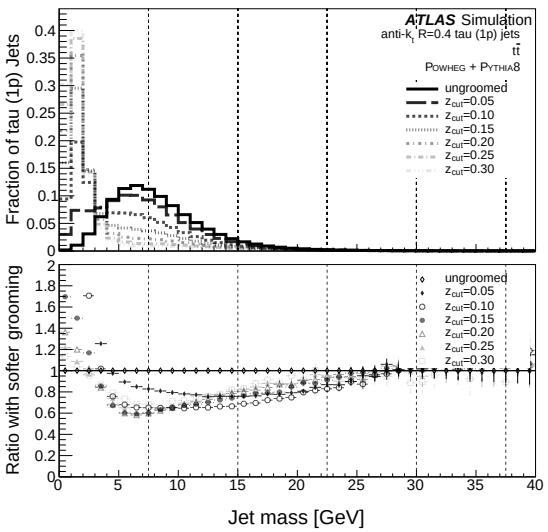
(b)



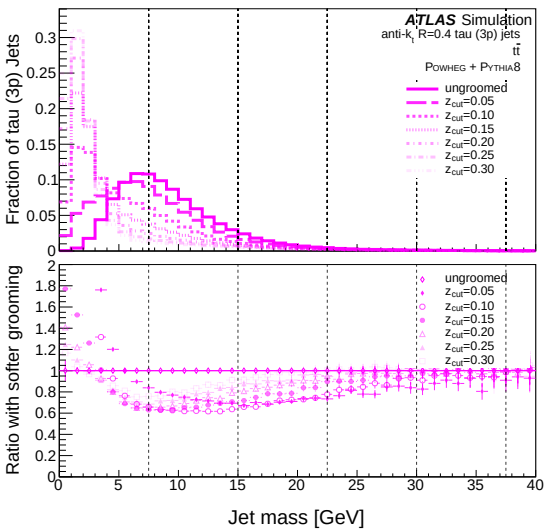
(c)



(d)



(e)



(f)

Figure A.3: Jet Mass at each strength of soft drop grooming for uds-, (a), gluon-, (b), b-, (c), c-, (d), 1-pronged tau -, (e) and 3-pronged tau -jets, (f)

SOFT DROP GROOMING ON TAU-JETS.

B

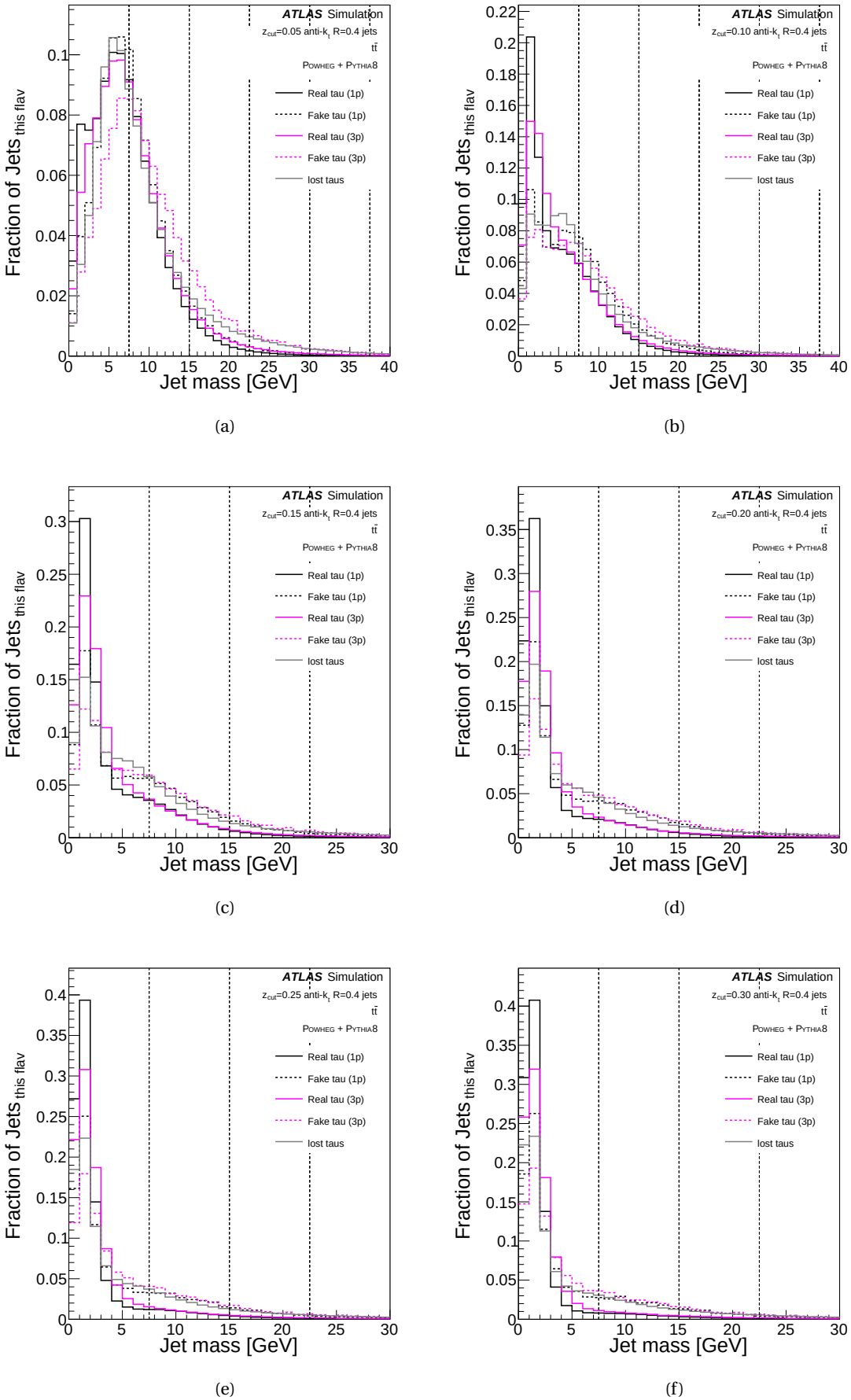


Figure B.1: Jet Mass for tau-flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

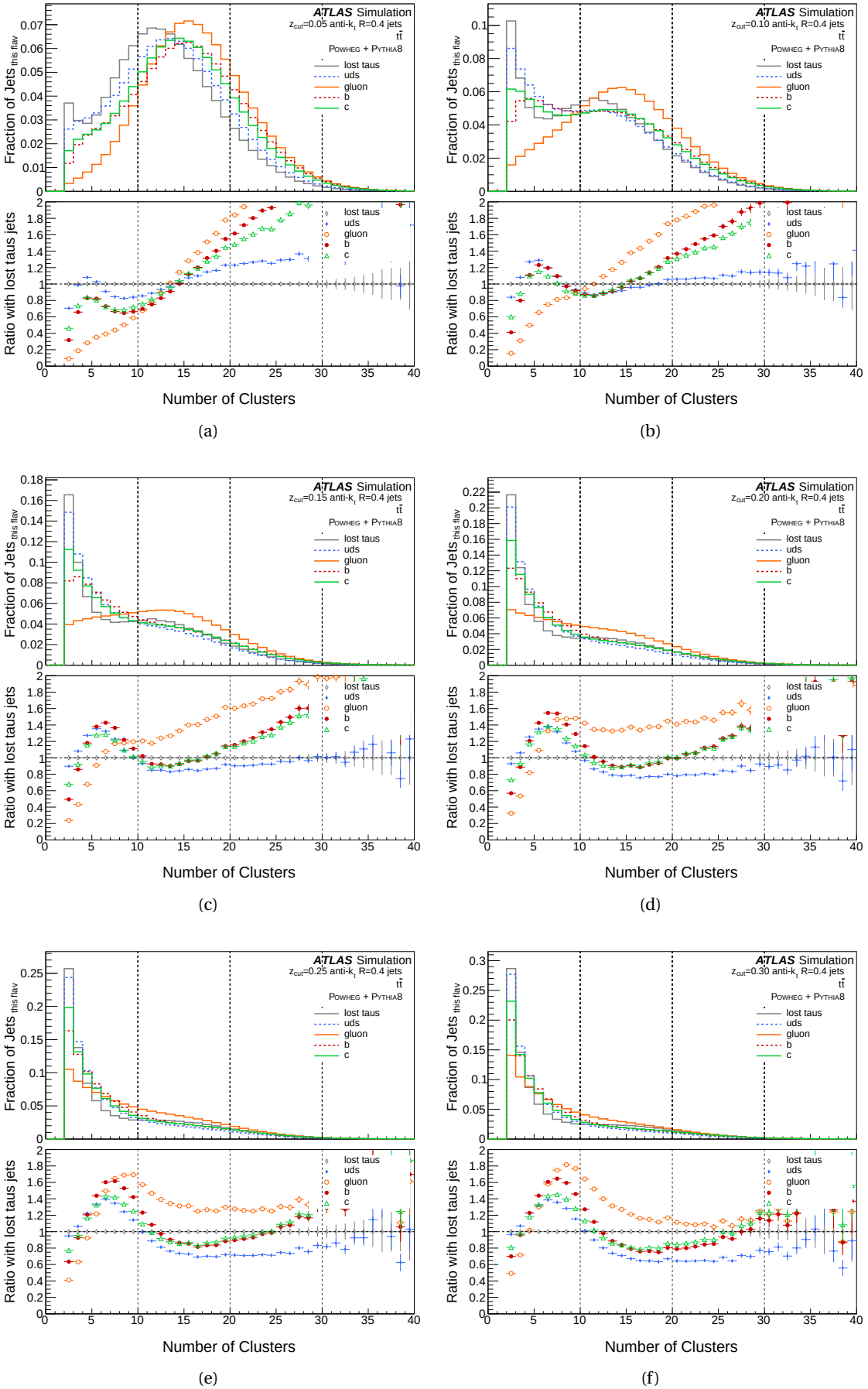


Figure B.2: Number of Clusters for all jets, including lost taus, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

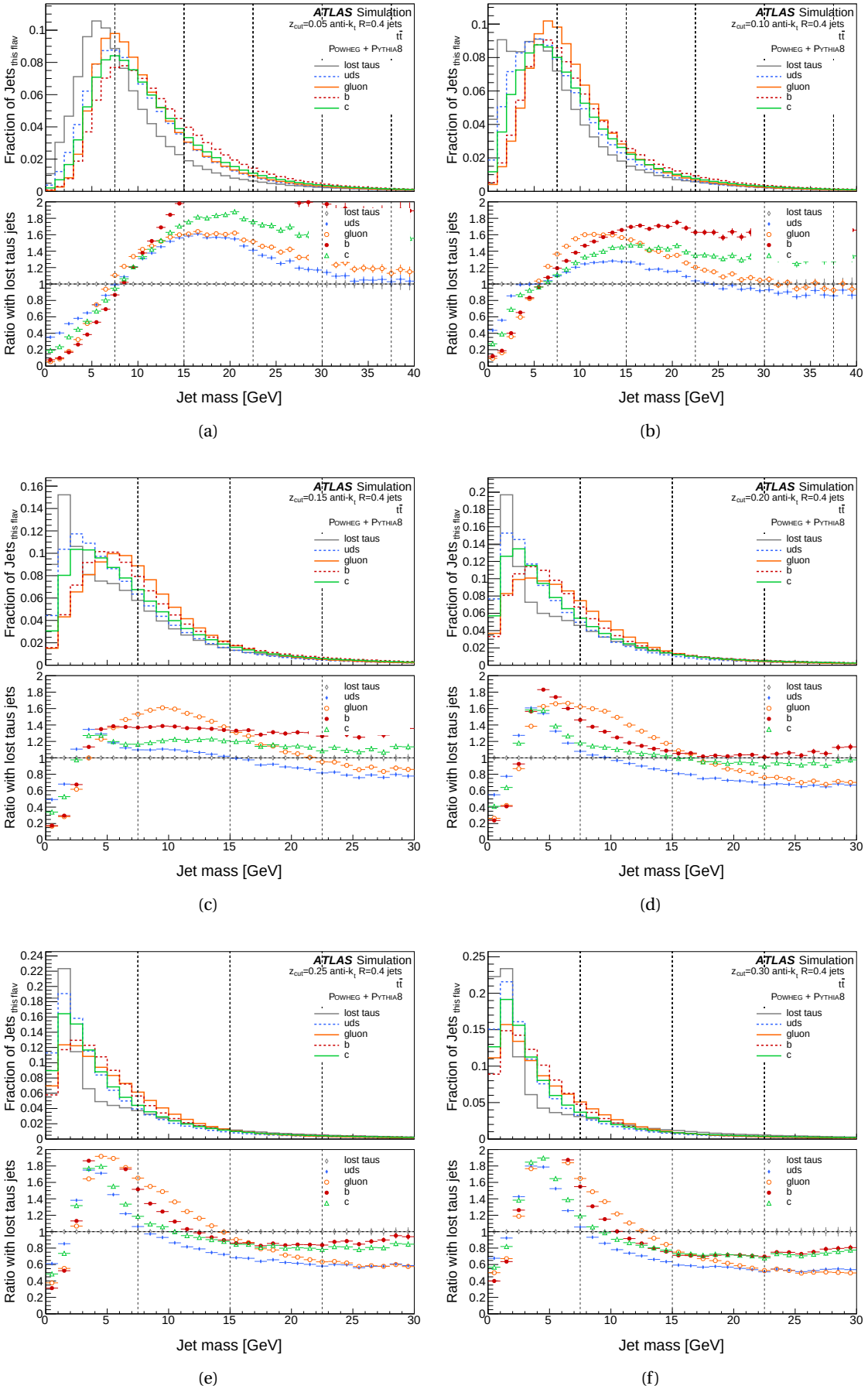
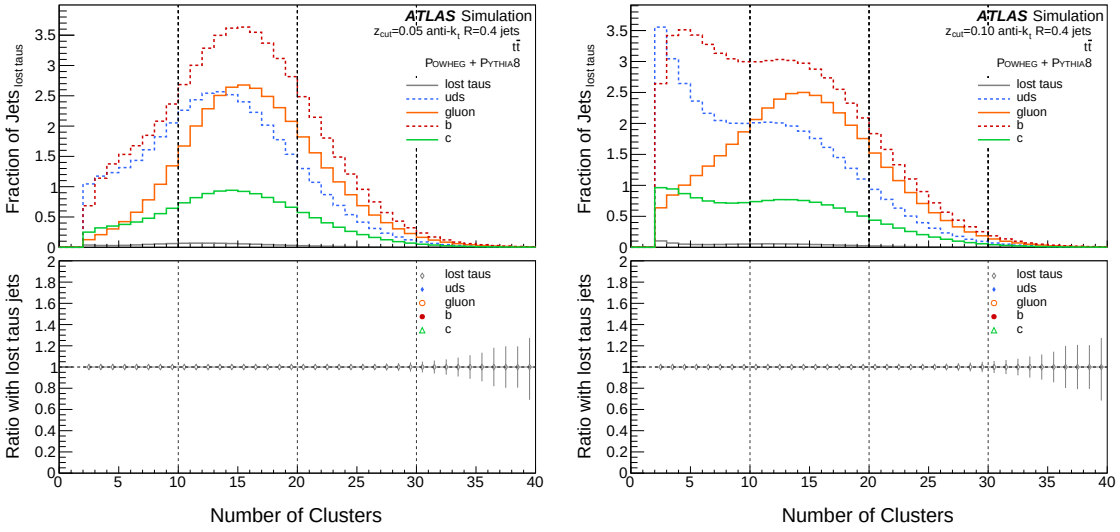
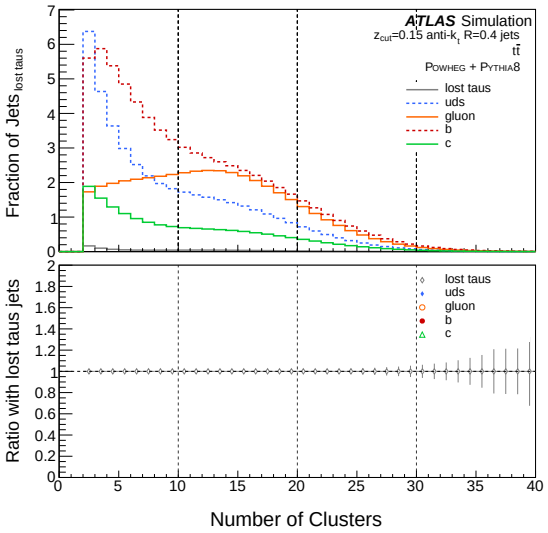


Figure B.3: Jet Mass for all flavour jets, including lost taus, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

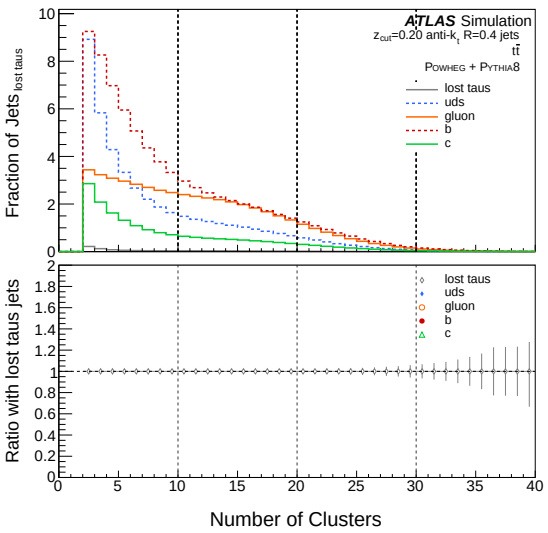


(a)

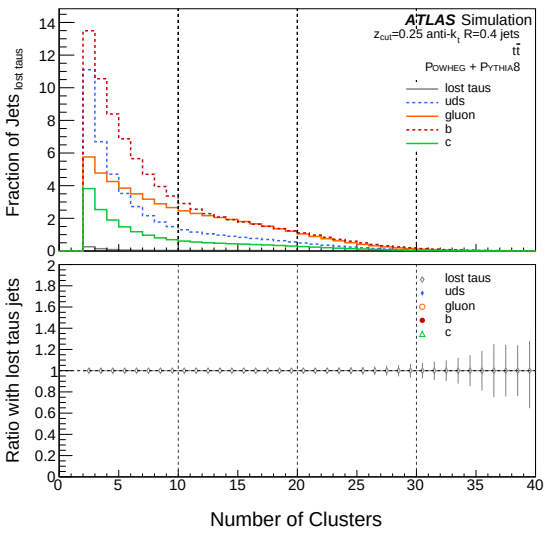
(b)



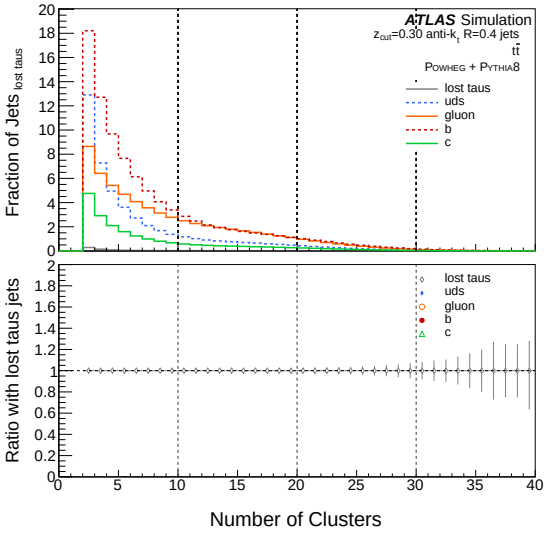
(c)



(d)



(e)



(f)

Figure B.4: Number of Clusters for all jets, including lost taus, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

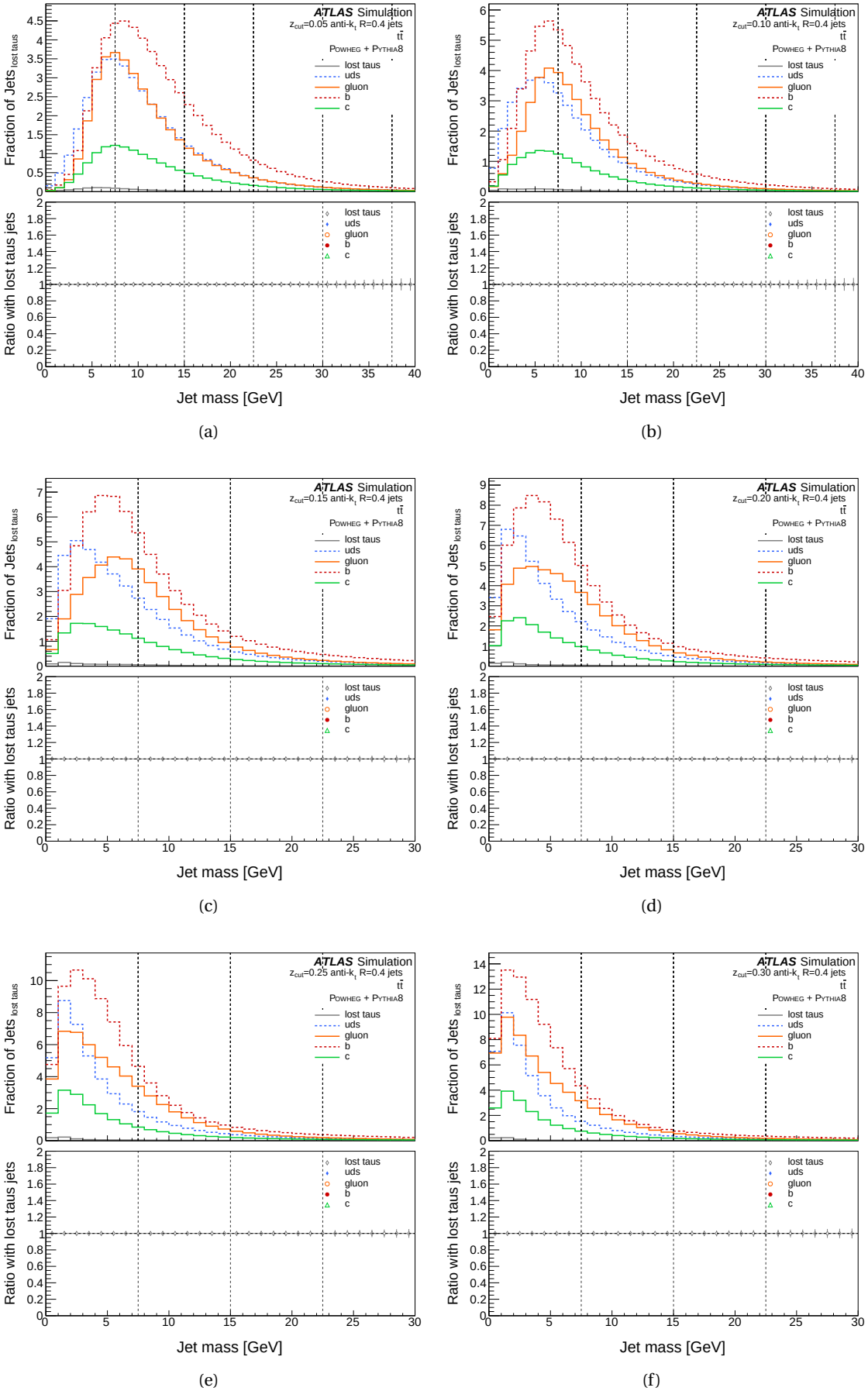
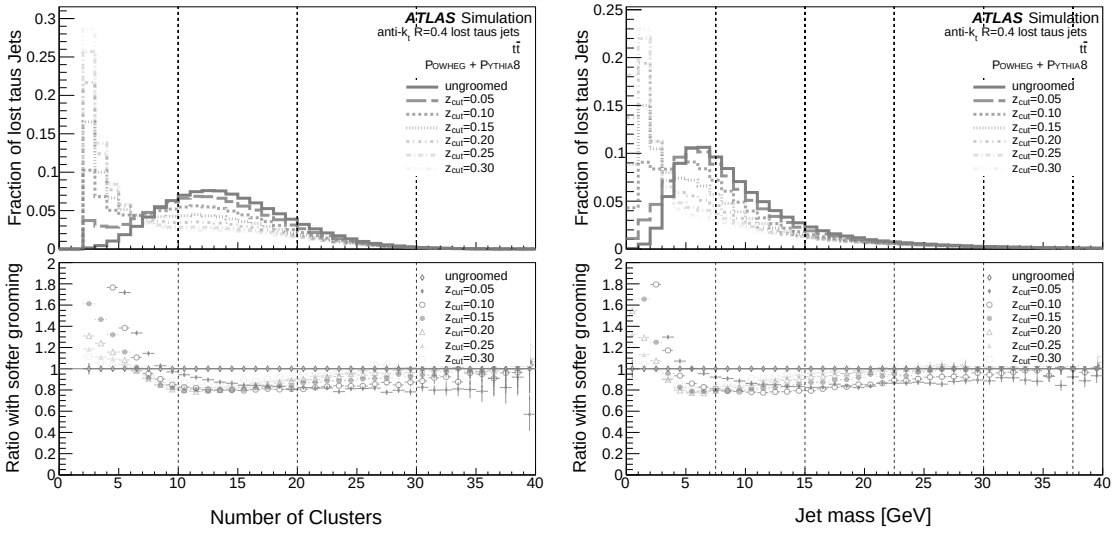


Figure B.5: Jet Mass for all flavour jets, including lost taus, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)



(a) (b)

Figure B.6: Jet Mass at each strength of soft drop grooming for lost taus

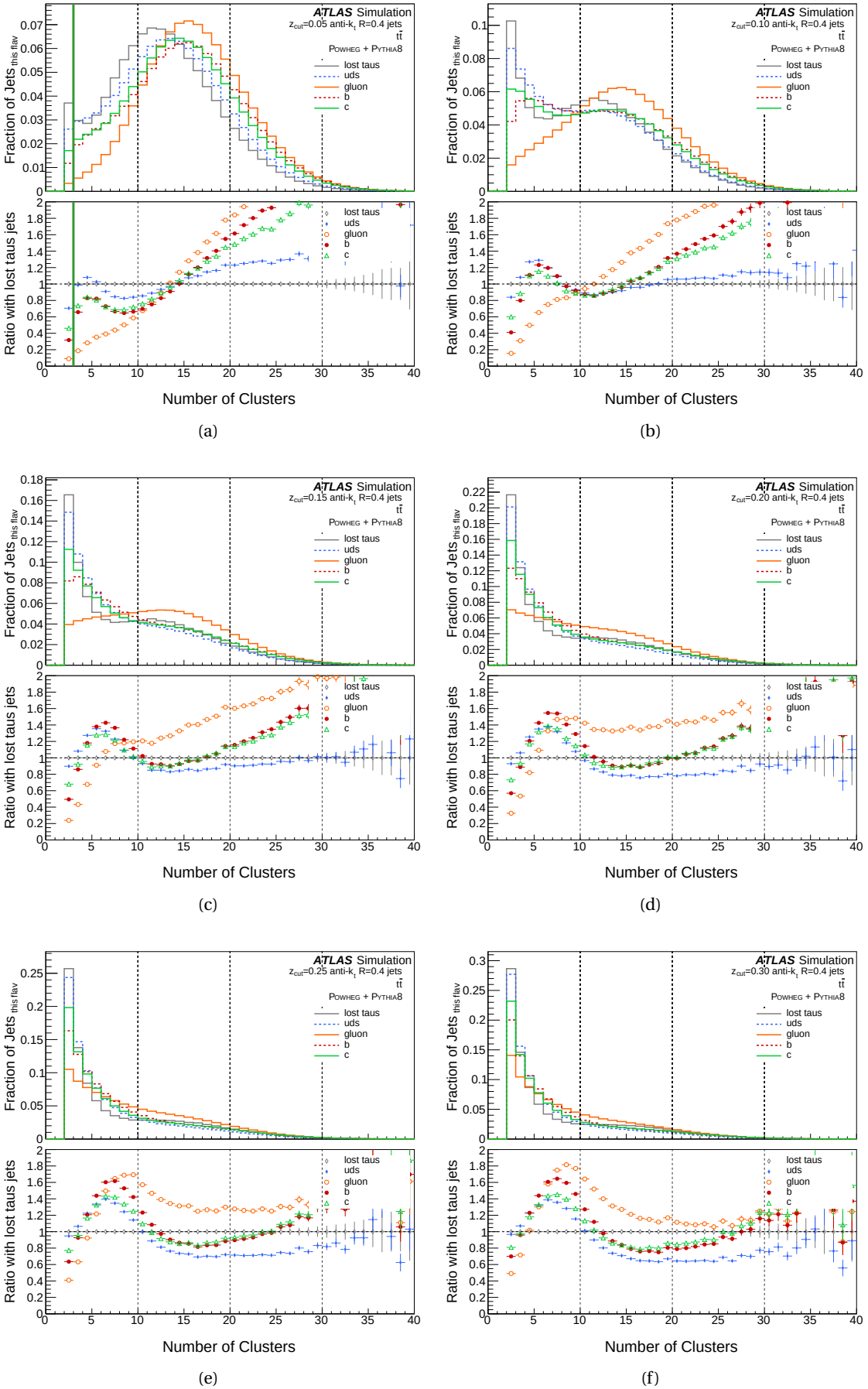


Figure B.7: Number of Clusters for all jets, including lost taus, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

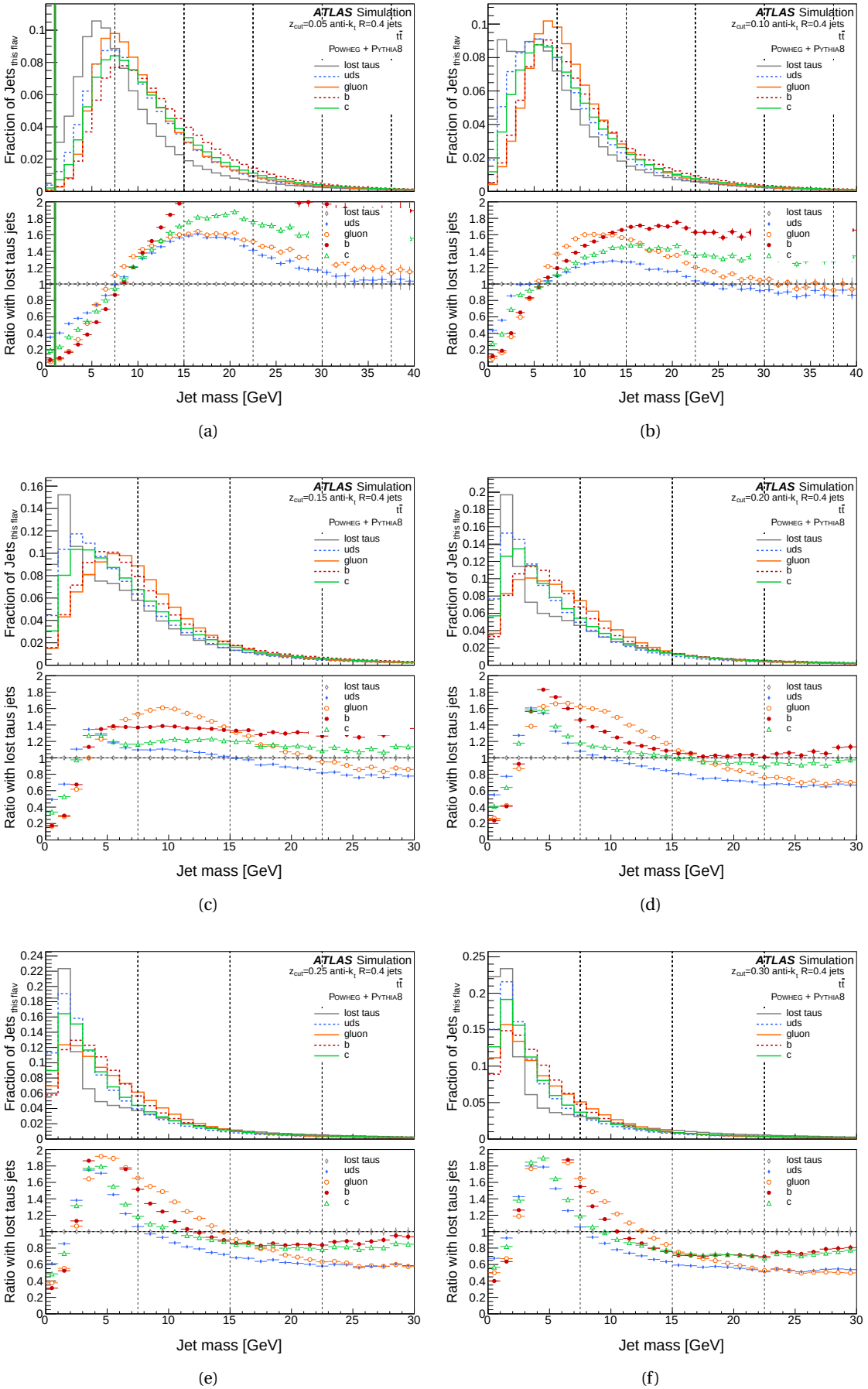
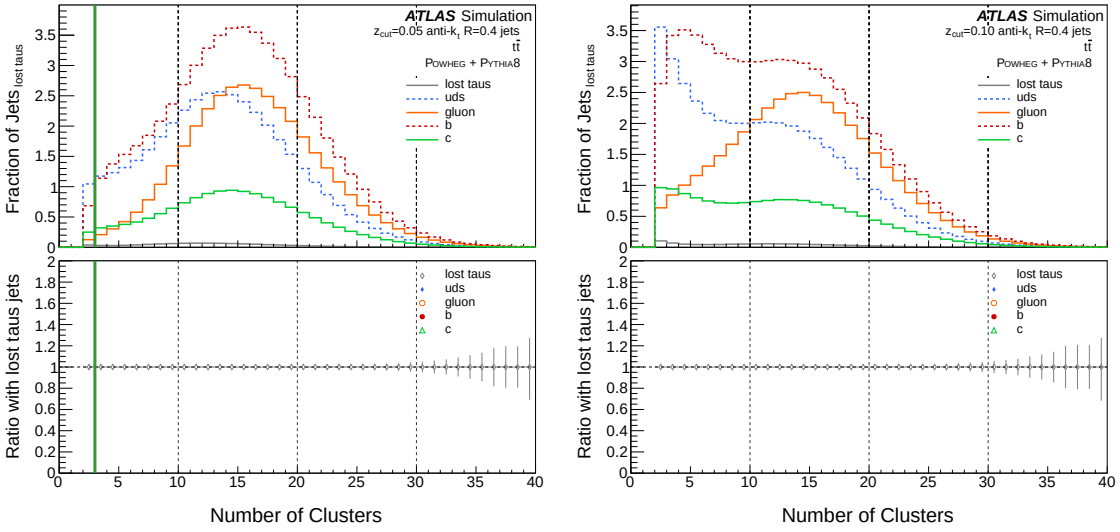
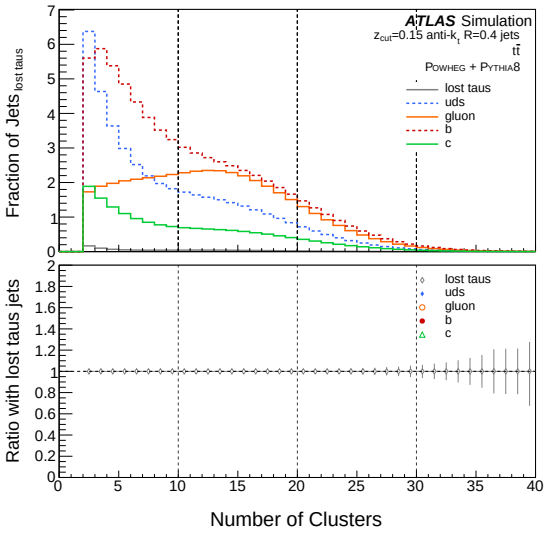


Figure B.8: Jet Mass for all flavour jets, including lost taus, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

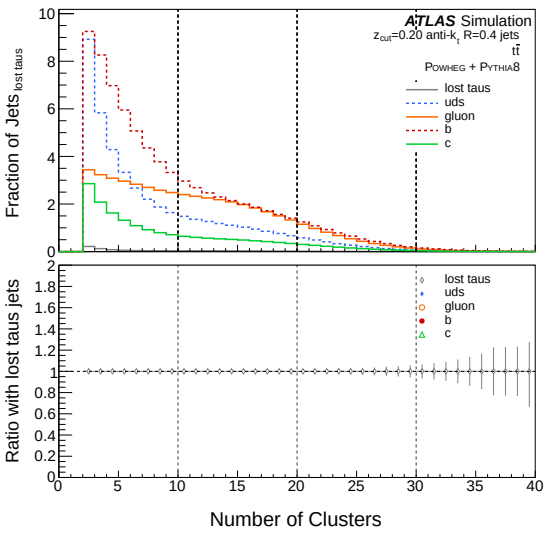


(a)

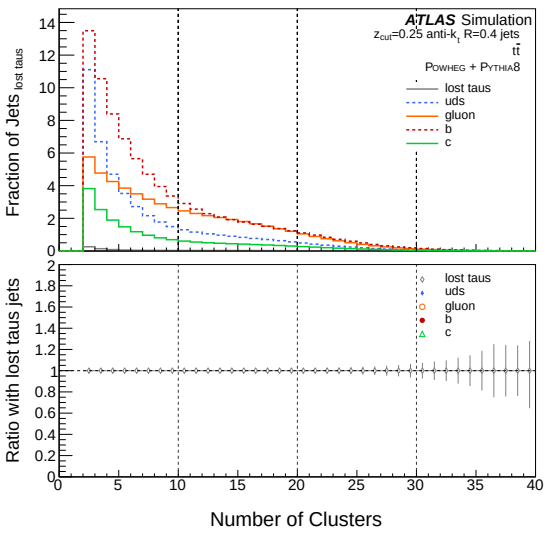
(b)



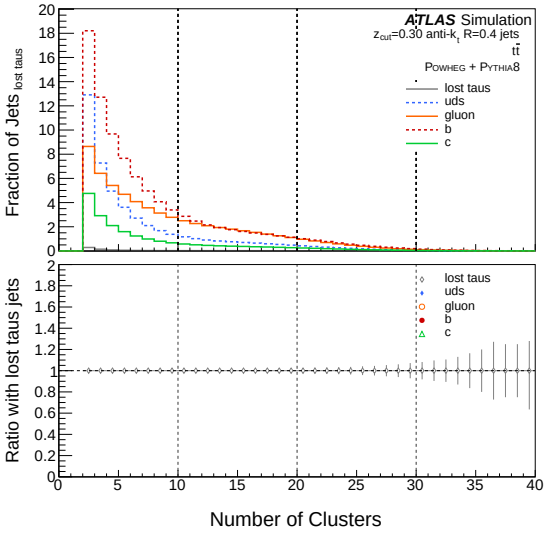
(c)



(d)

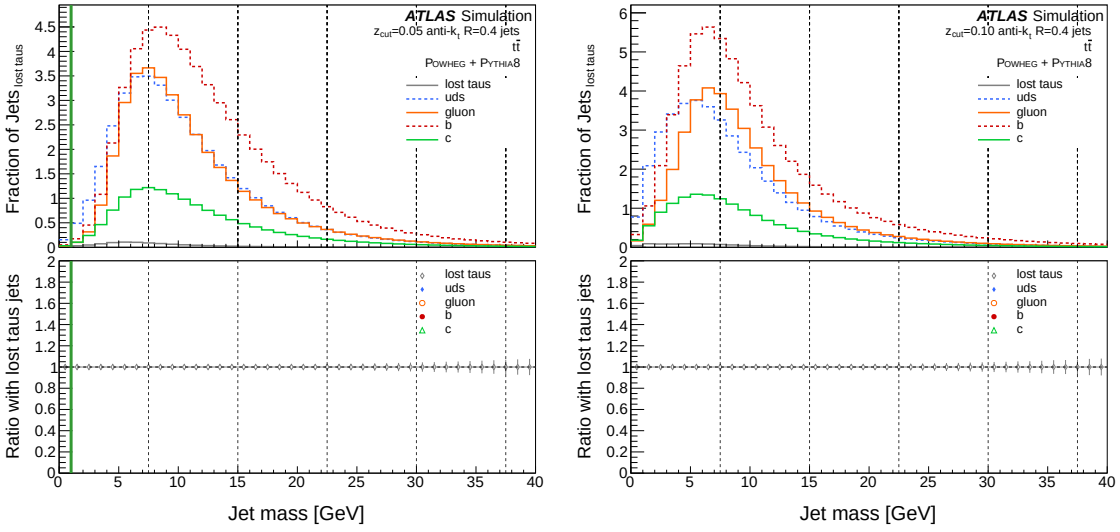


(e)



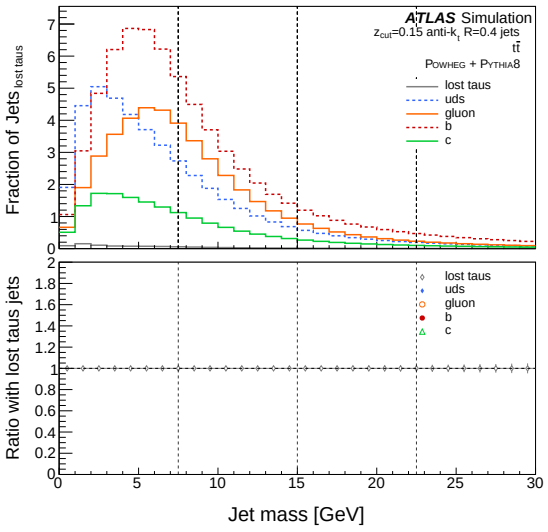
(f)

Figure B.9: Number of Clusters for all jets, including lost taus, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

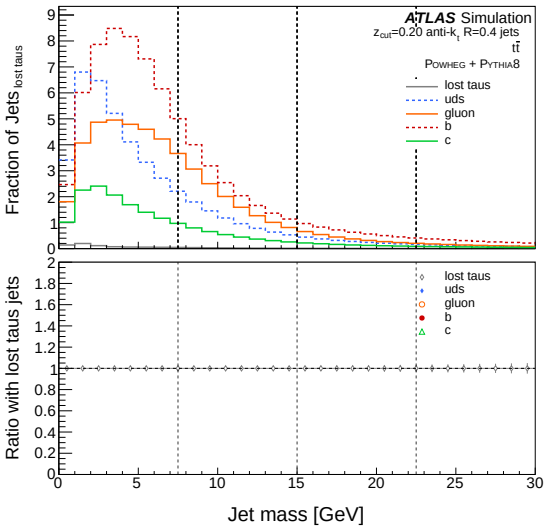


(a)

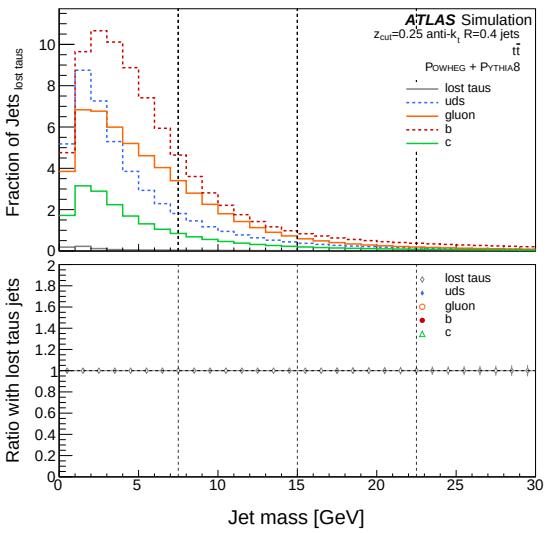
(b)



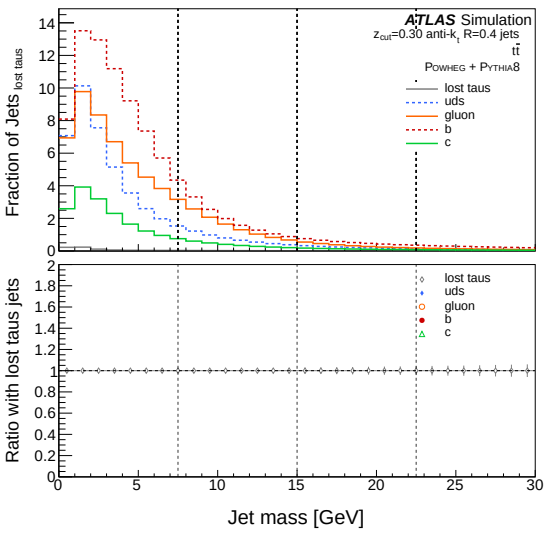
(c)



(d)



(e)



(f)

Figure B.10: Jet Mass for all flavour jets, including lost taus, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f)

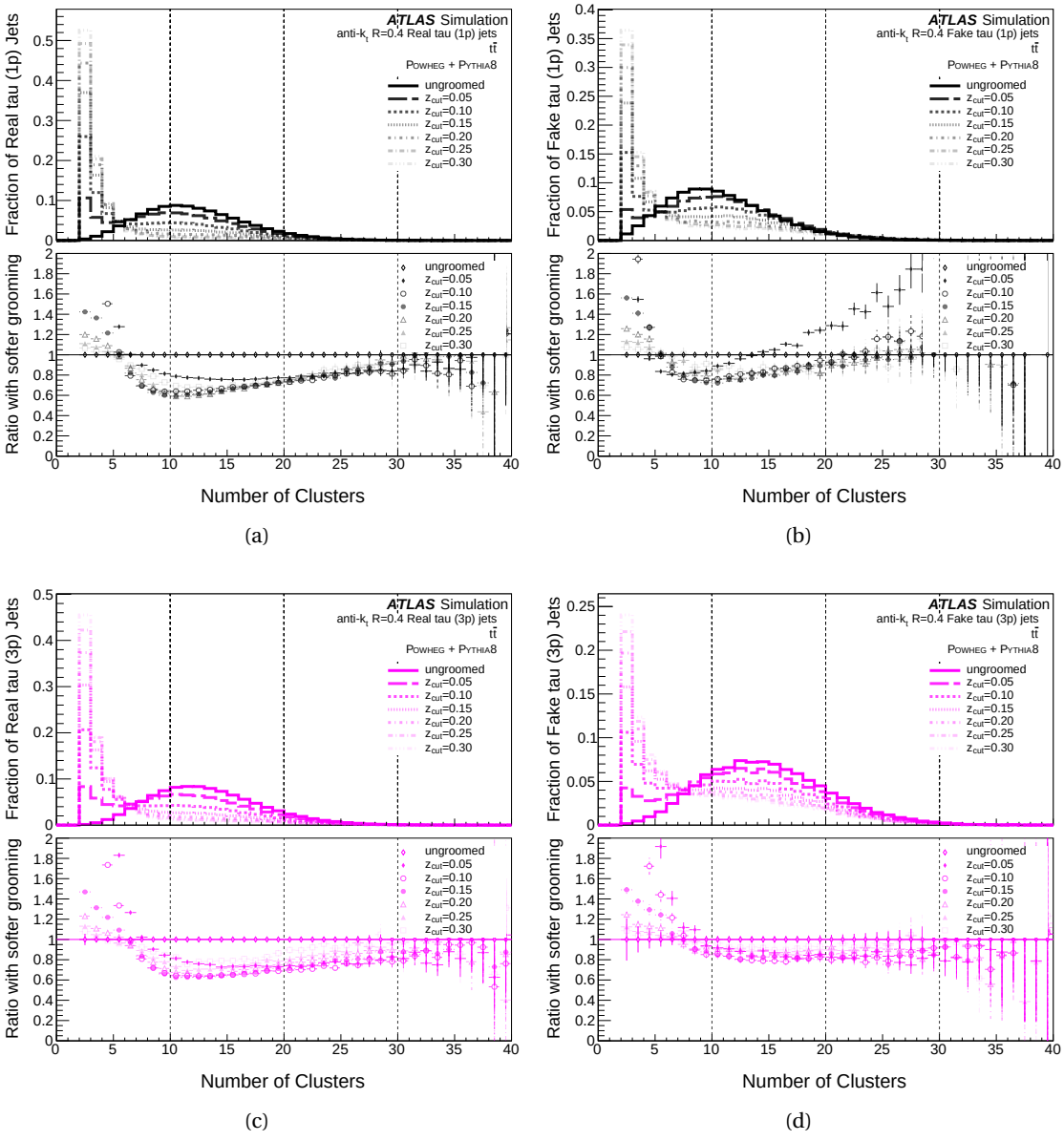
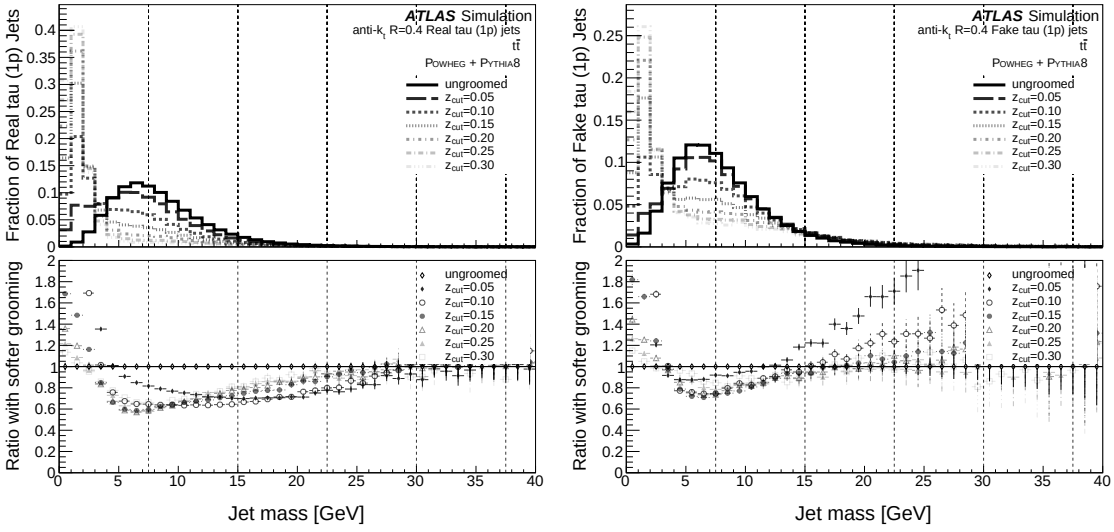
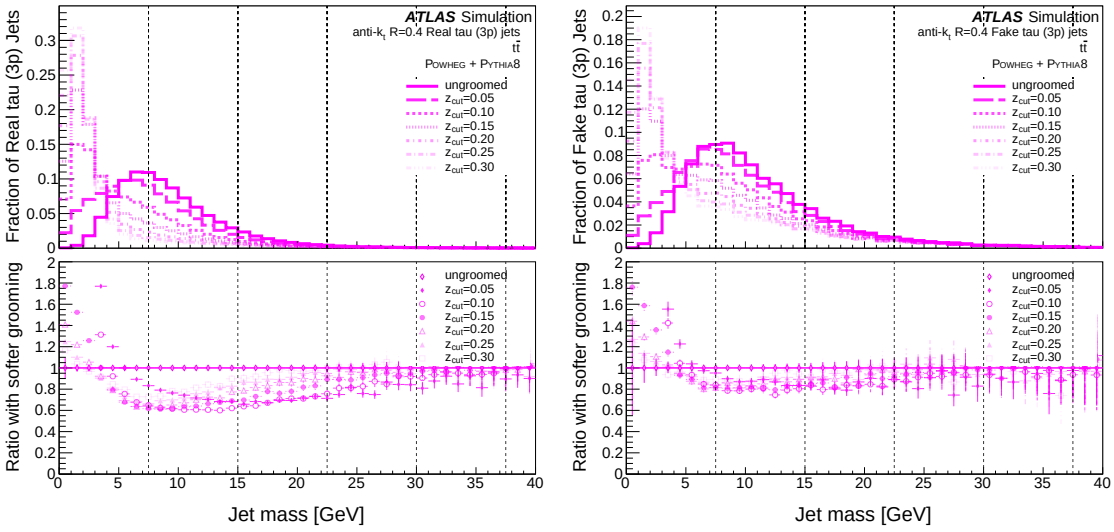


Figure B.11: Number of Clusters at each strength of soft drop grooming for real and fake taus

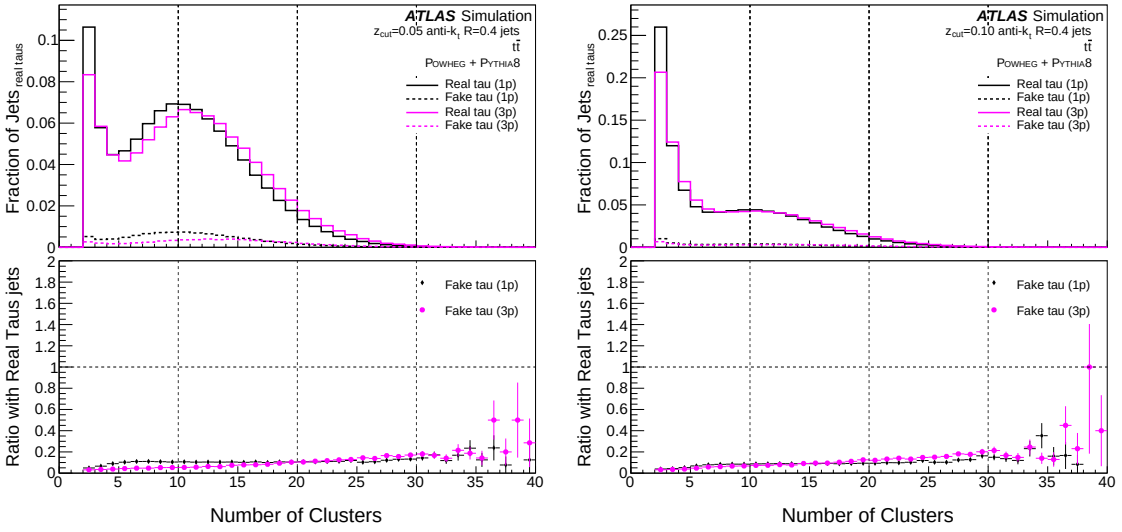


(a) (b)



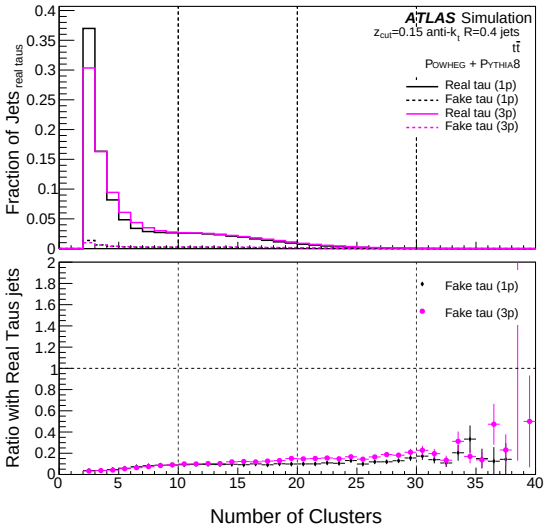
(c) (d)

Figure B.12: Jet Mass at each strength of soft drop grooming for real and fake taus

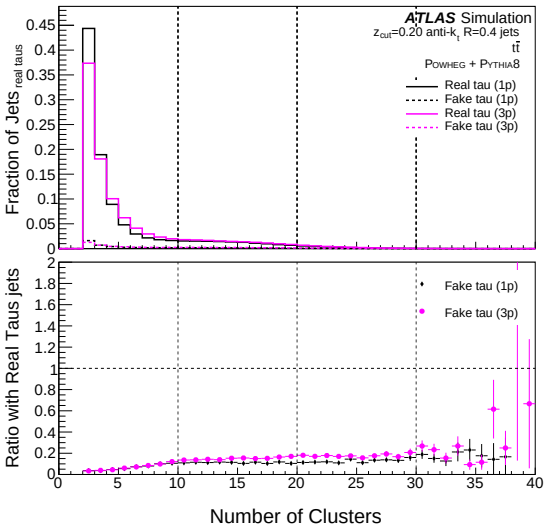


(a)

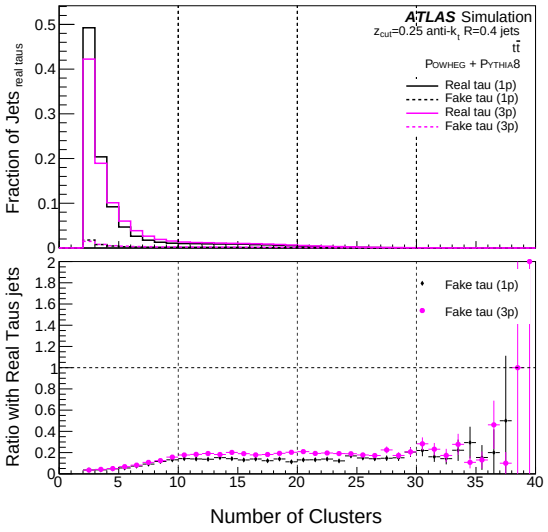
(b)



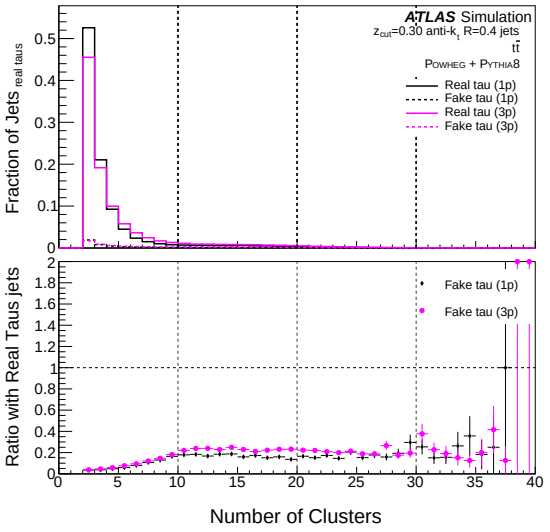
(c)



(d)



(e)



(f)

Figure B.13: Number of Clusters for tau-flavour jets, groomed with strength Z_{cut} = 0.05, (a), Z_{cut} = 0.10, (b), Z_{cut} = 0.15, (c), Z_{cut} = 0.20, (d), Z_{cut} = 0.25, (e), Z_{cut} = 0.30, (f), all normalised to the number of real taus for each prong value.

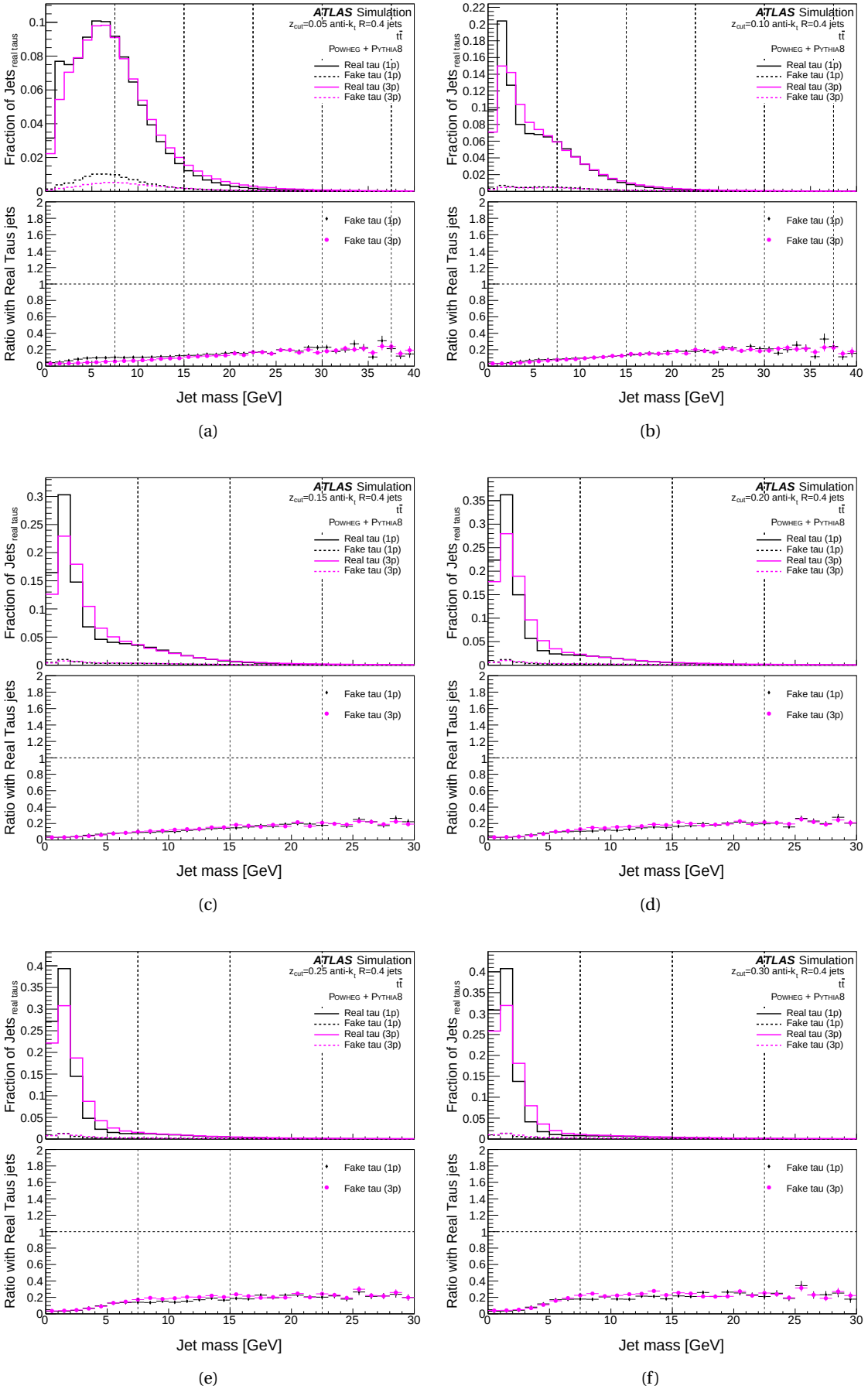
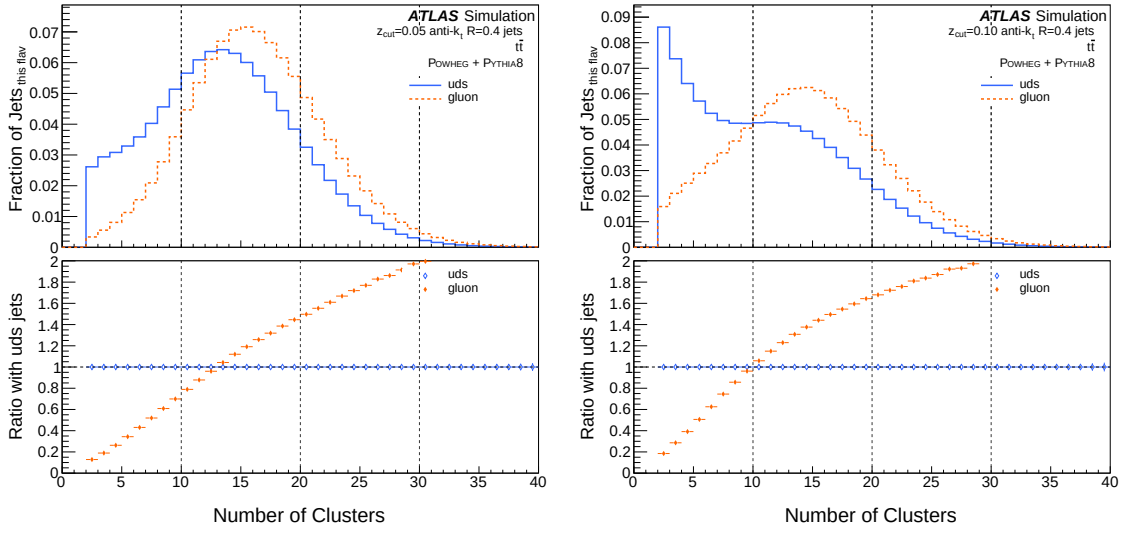


Figure B.14: Jet Mass for tau-flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of real taus for each prong value.

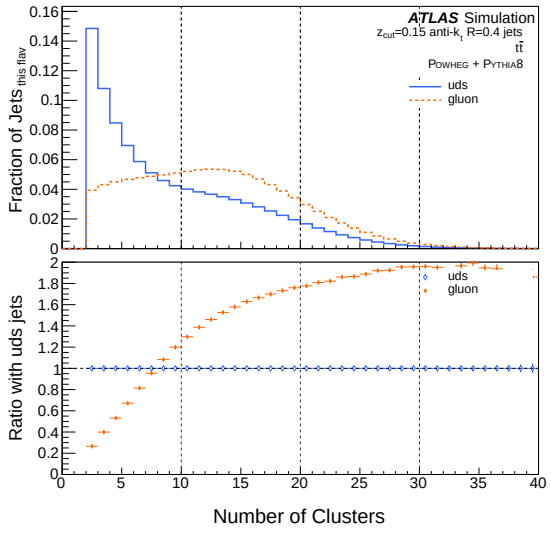
C

Soft drop grooming on light-flavoured flavour-tagged small jets().

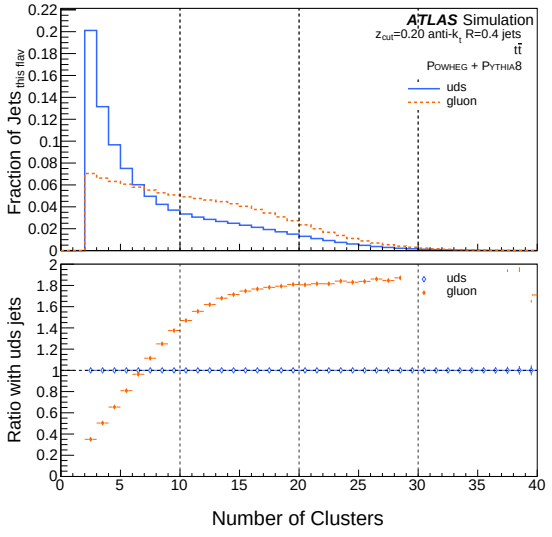


(a)

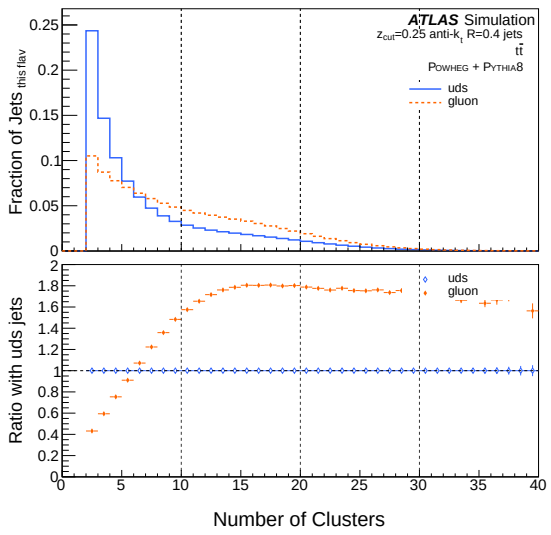
(b)



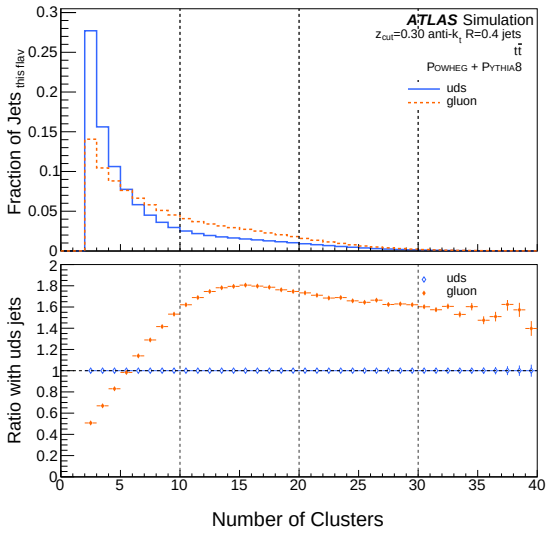
(c)



(d)



(e)



(f)

Figure C.1: Number of Clusters for quark- and gluon-flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of uds jets.

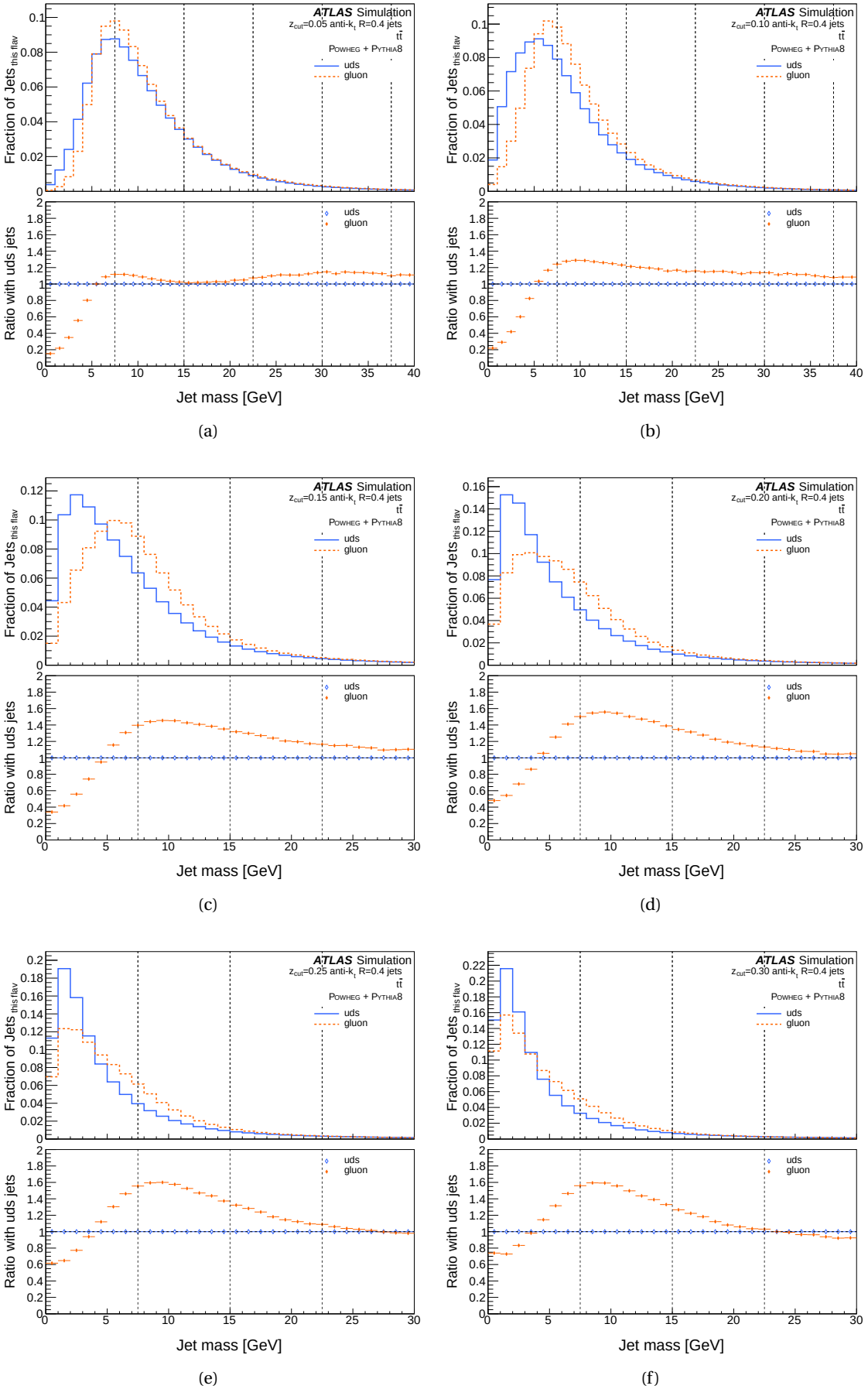
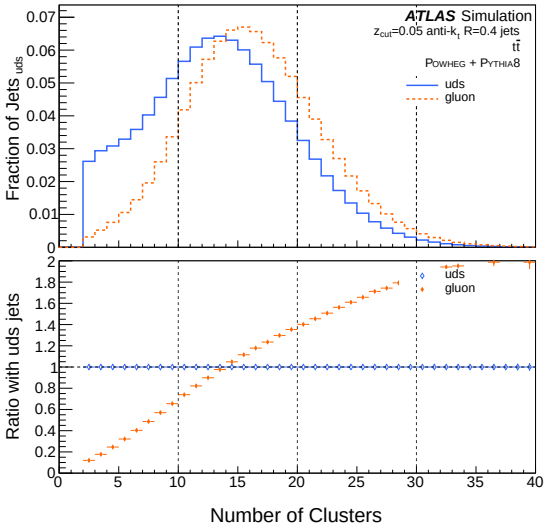
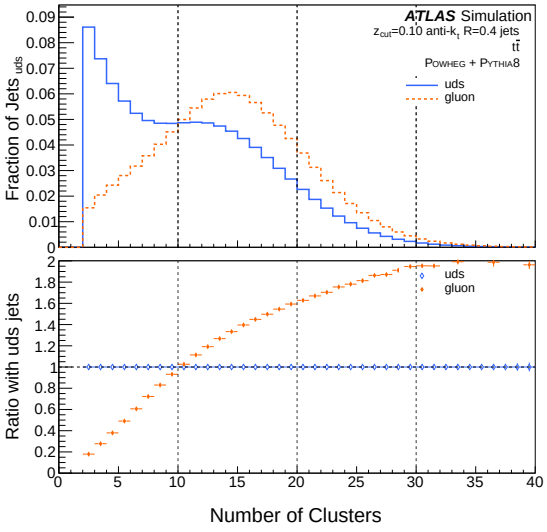


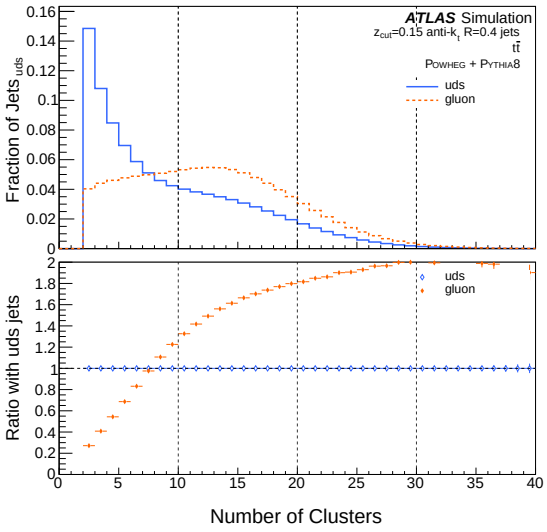
Figure C.2: Jet Mass for quark- and gluon-flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of uds jets.



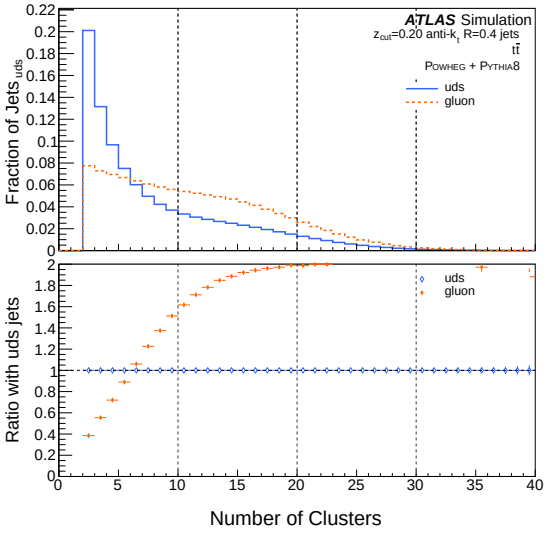
(a)



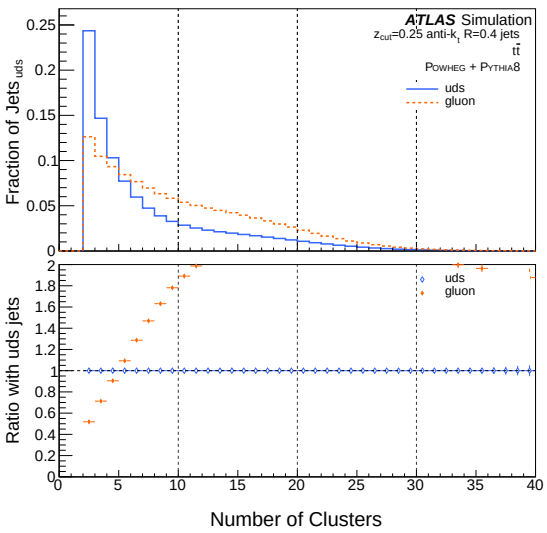
(b)



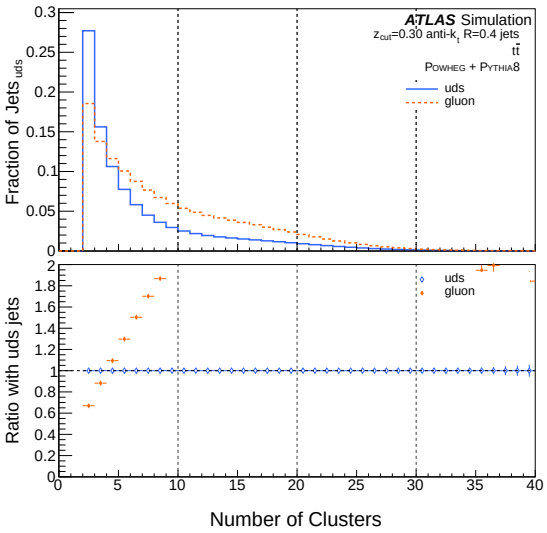
(c)



(d)



(e)



(f)

Figure C.3: Number of Clusters for quark- and gluon-flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of uds jets.

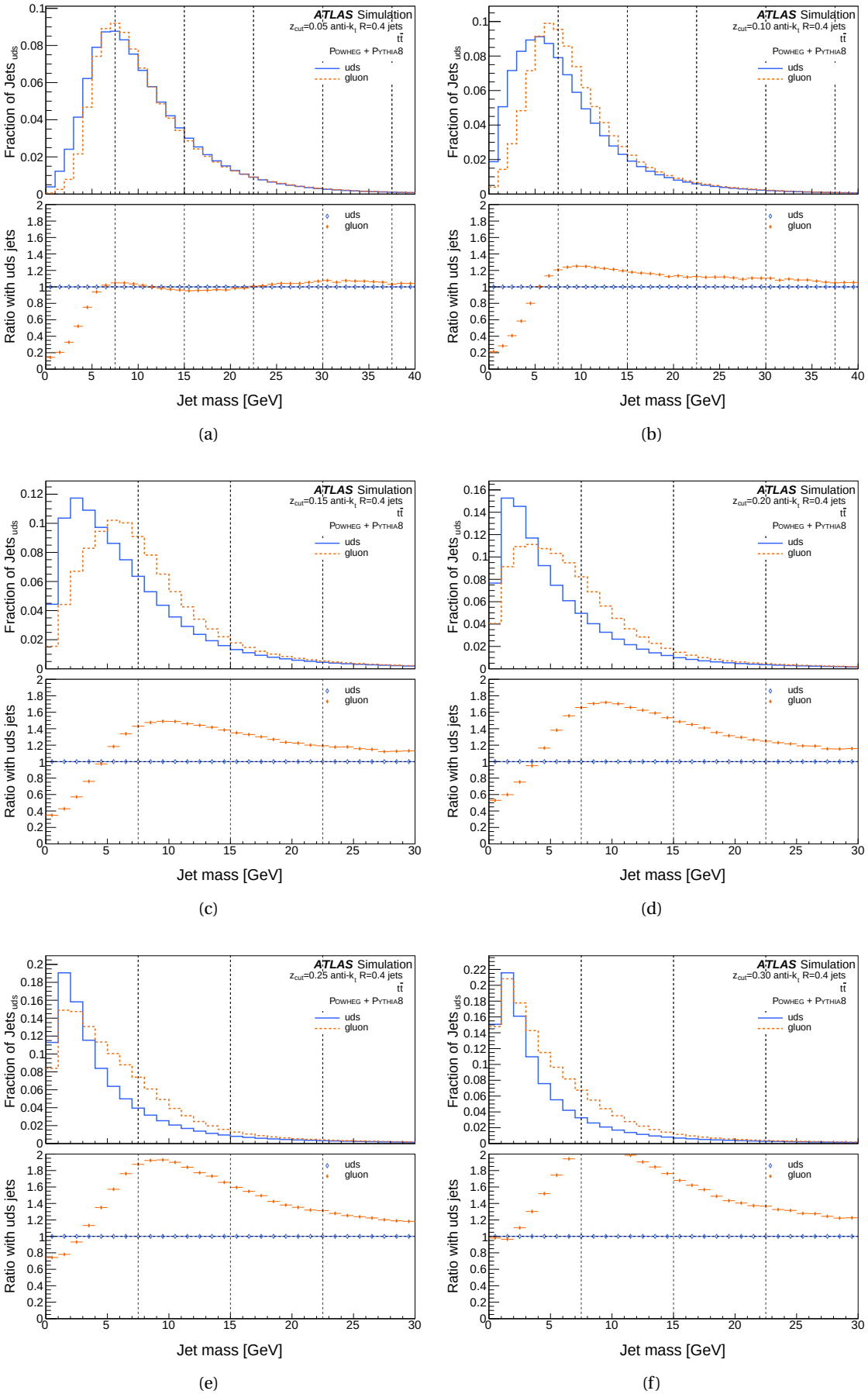


Figure C.4: Jet Mass for quark- and gluon-flavour jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of uds jets.

SOFT DROP GROOMING ON FLAVOUR-TAGGED SMALL JETS() WITH B-TAGGING

D

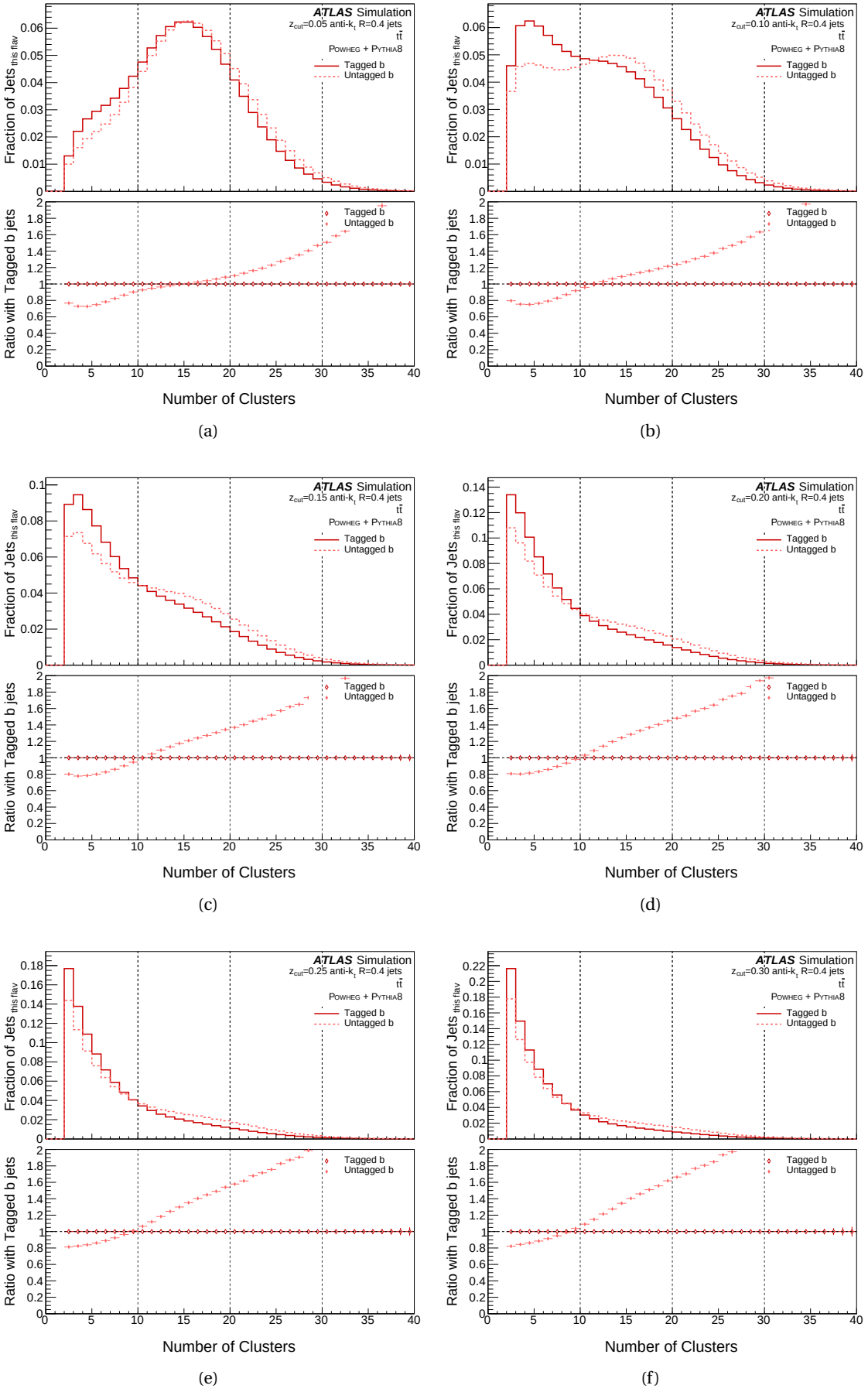
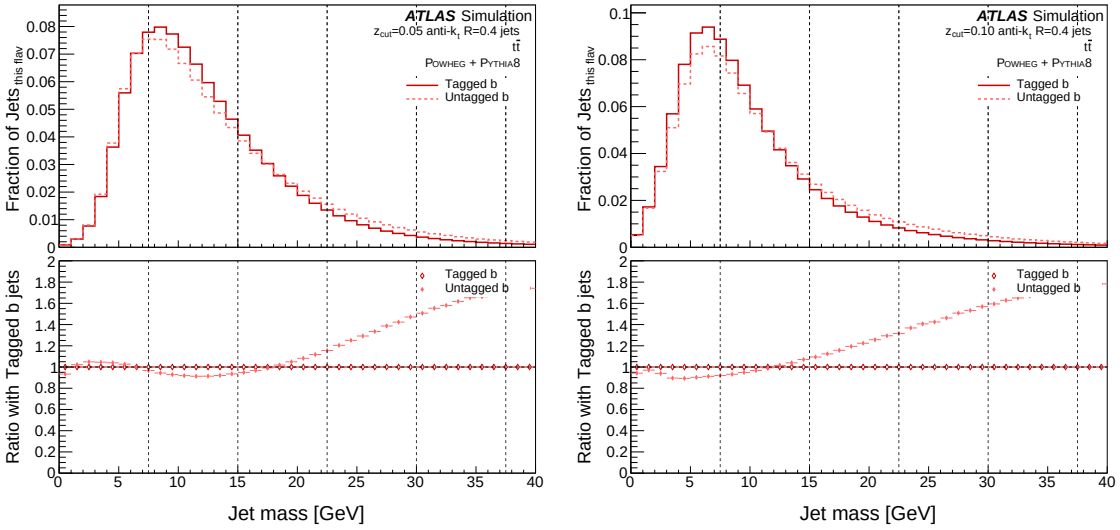
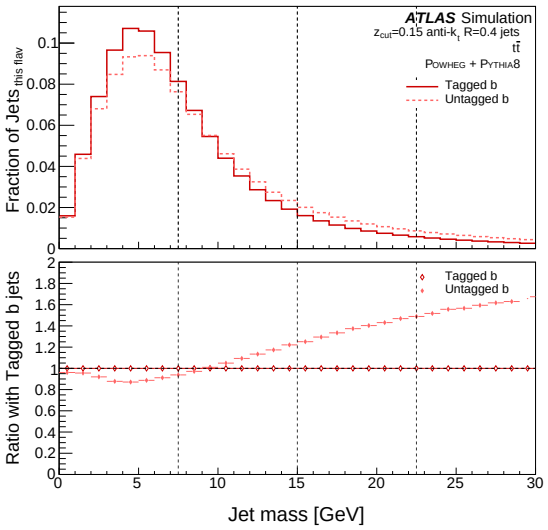


Figure D.1: Number of Clusters for tagged and untagged *b*-jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of tagged b jets.

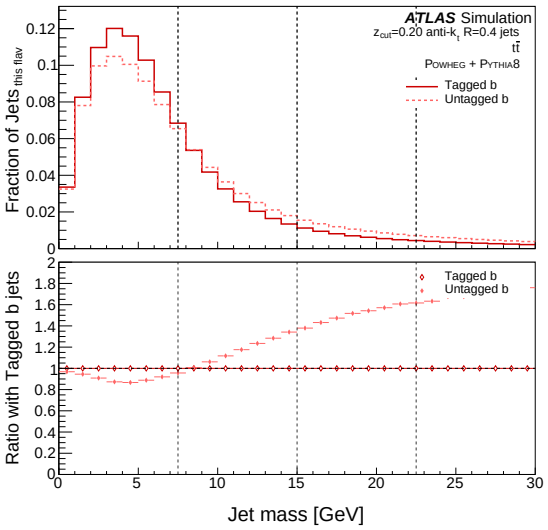


(a)

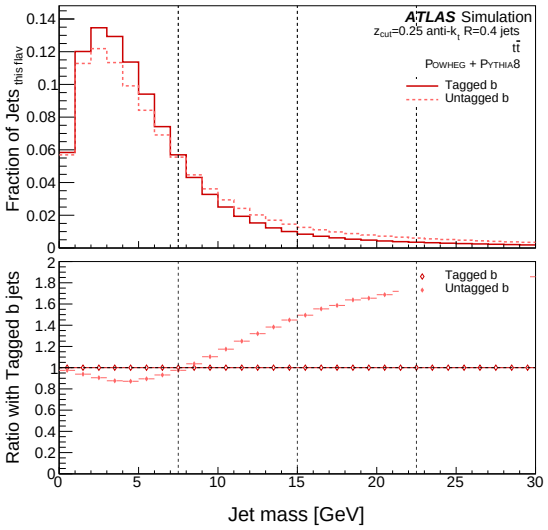
(b)



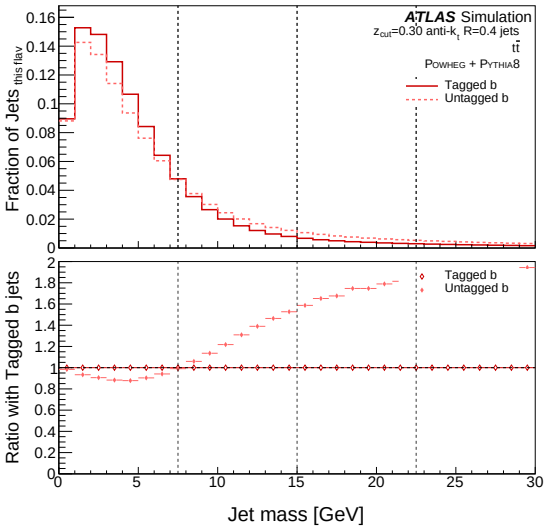
(c)



(d)

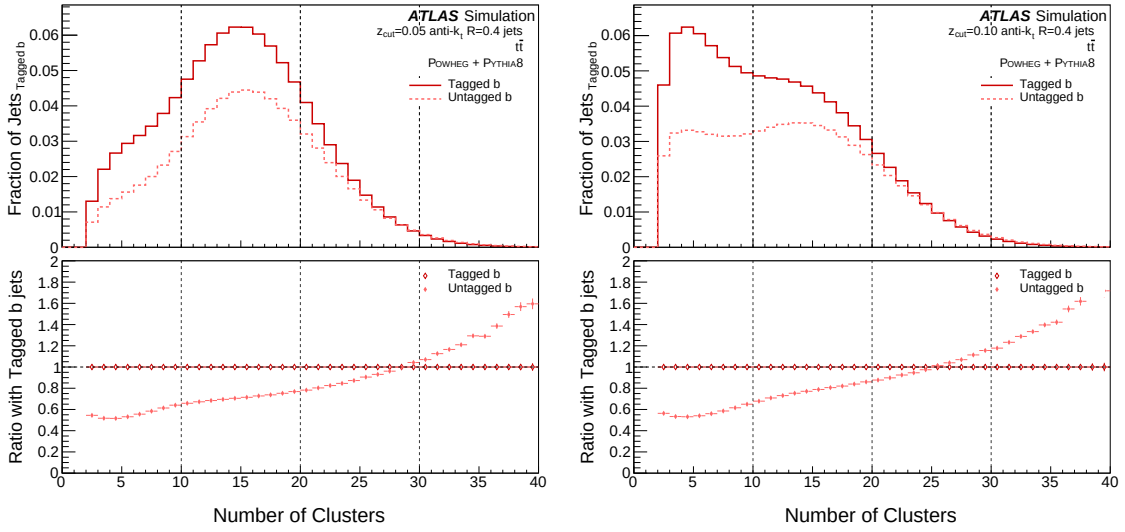


(e)



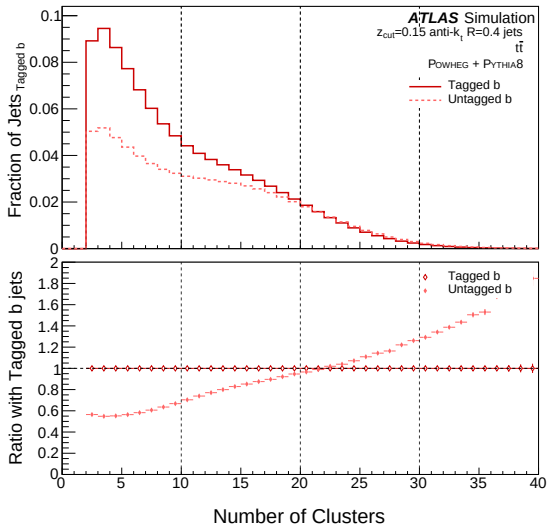
(f)

Figure D.2: Jet Mass for tagged and untagged *b*-jets, groomed with strength Z_{cut} = 0.05, (a), Z_{cut} = 0.10, (b), Z_{cut} = 0.15, (c), Z_{cut} = 0.20, (d), Z_{cut} = 0.25, (e), Z_{cut} = 0.30, (f), all normalised to the number of tagged b jets.

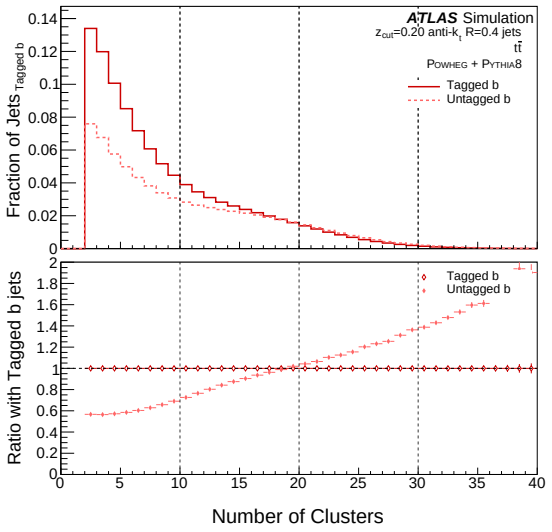


(a)

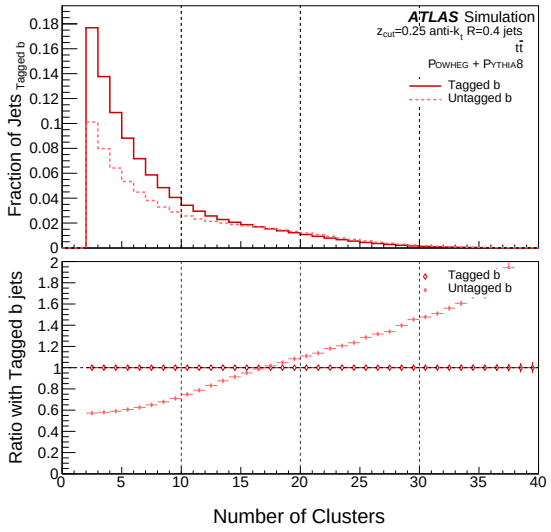
(b)



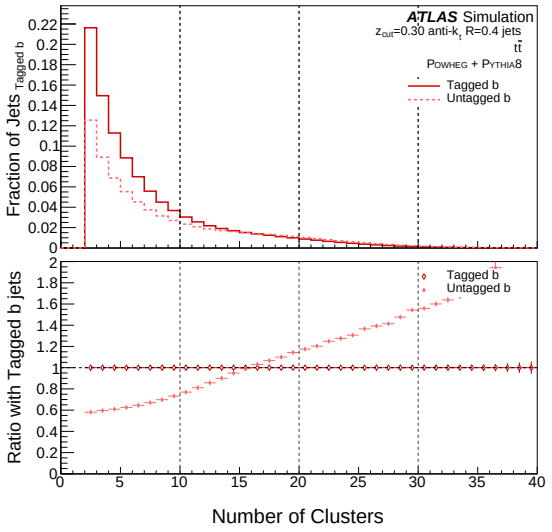
(c)



(d)

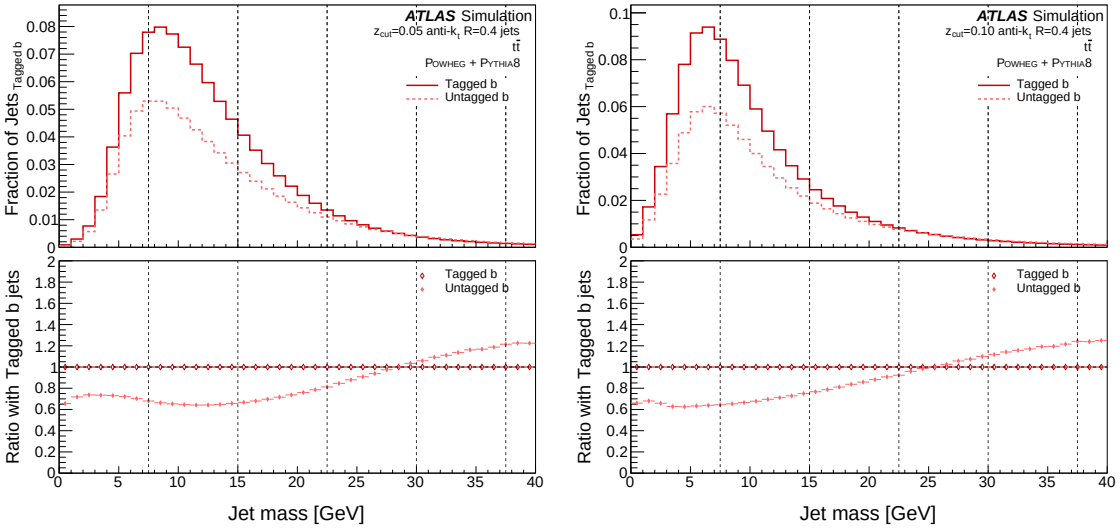


(e)



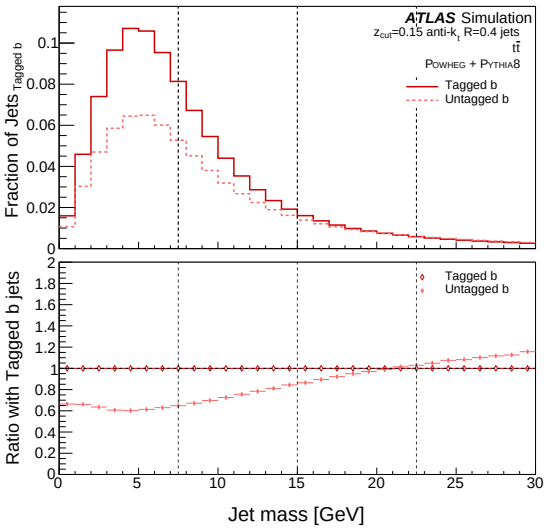
(f)

Figure D.3: Number of Clusters for tagged and untagged *b*-jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of tagged b jets.

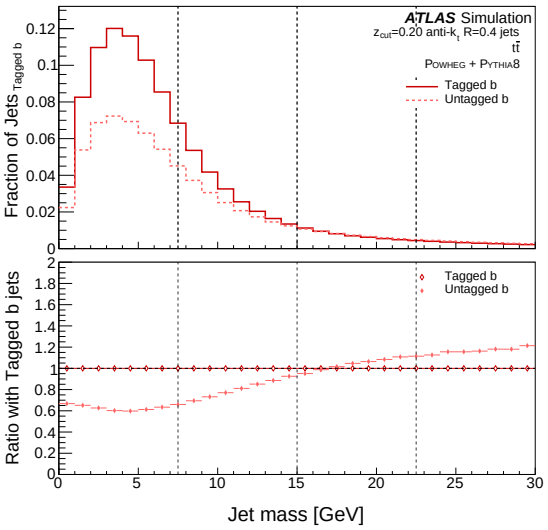


(a)

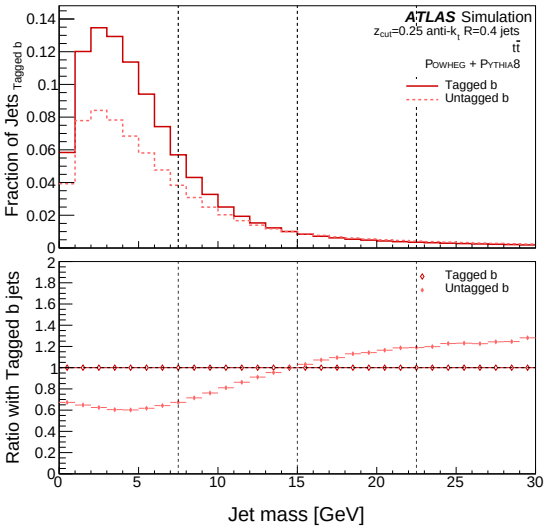
(b)



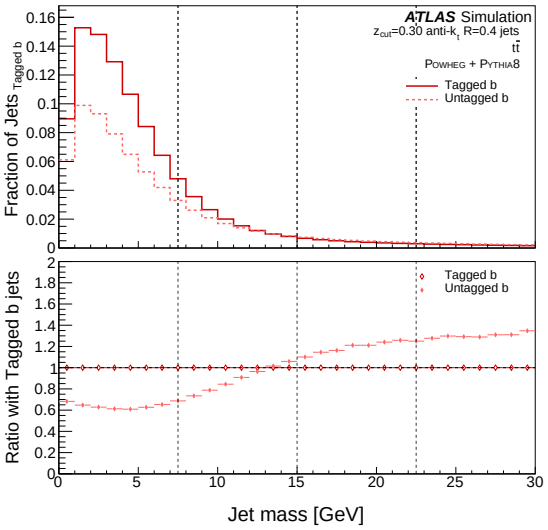
(c)



(d)



(e)



(f)

Figure D.4: Jet Mass for tagged and untagged *b*-jets, groomed with strength Z_{cut} = 0.05, (a), Z_{cut} = 0.10, (b), Z_{cut} = 0.15, (c), Z_{cut} = 0.20, (d), Z_{cut} = 0.25, (e), Z_{cut} = 0.30, (f), all normalised to the number of tagged b jets.

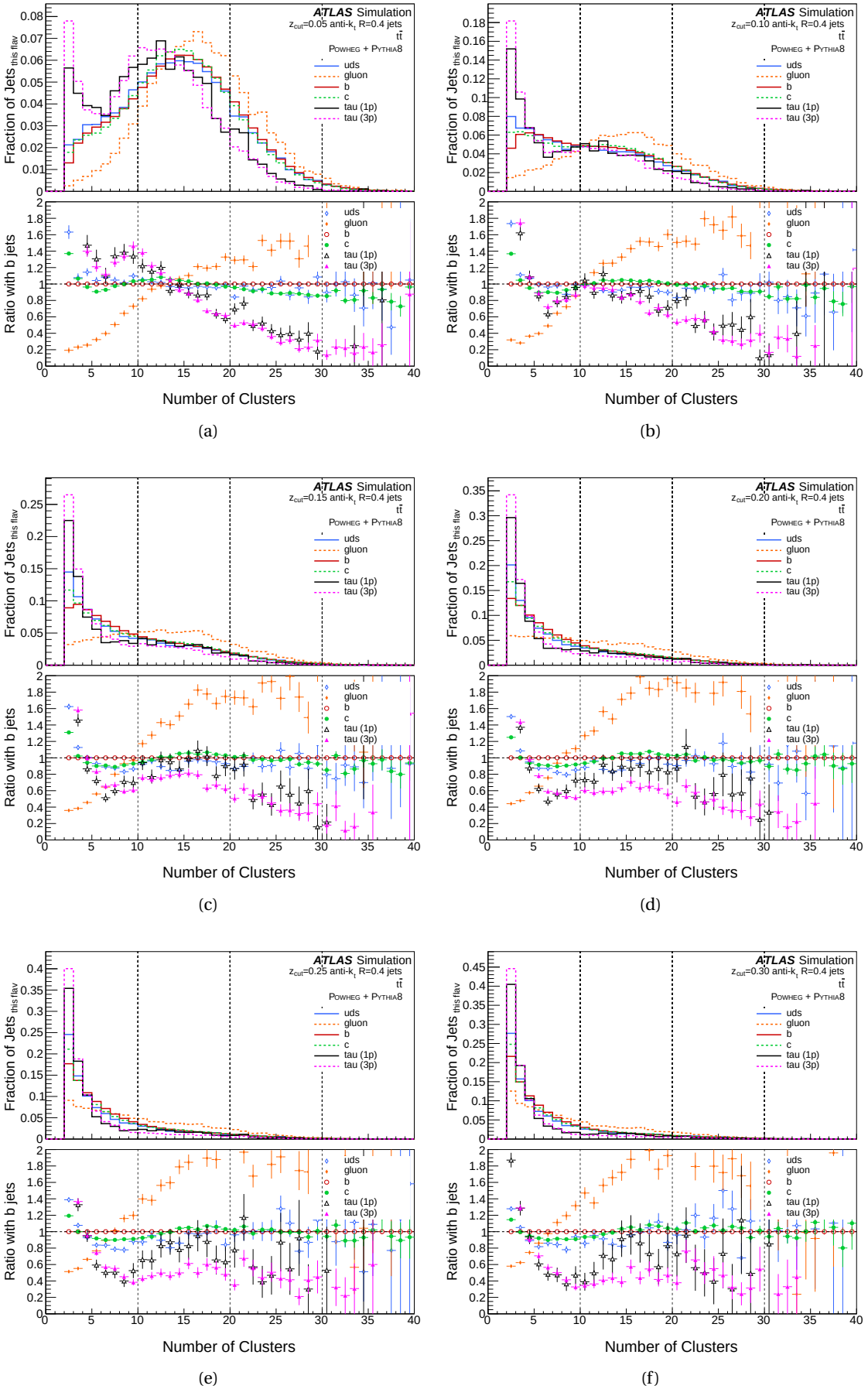
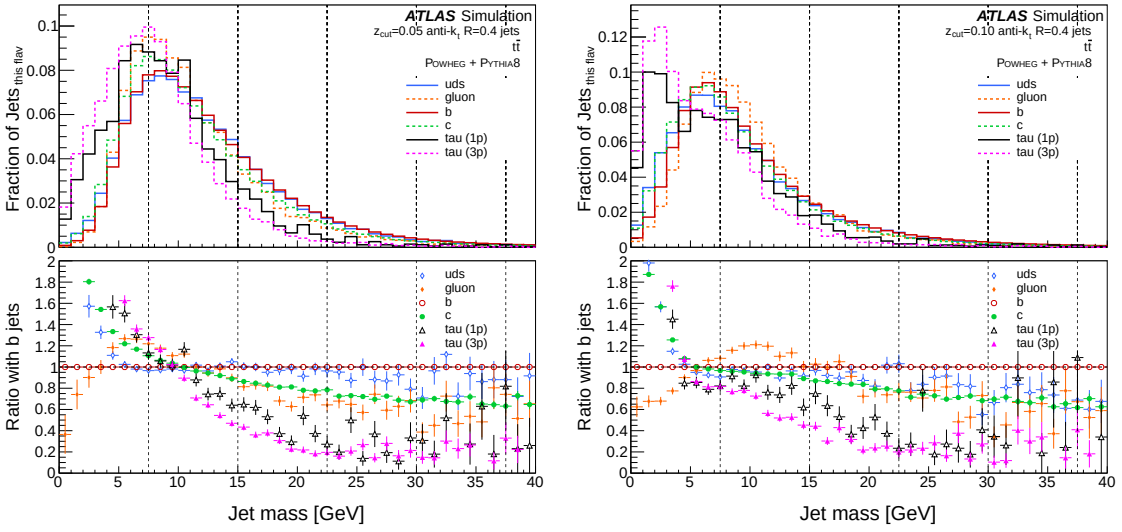
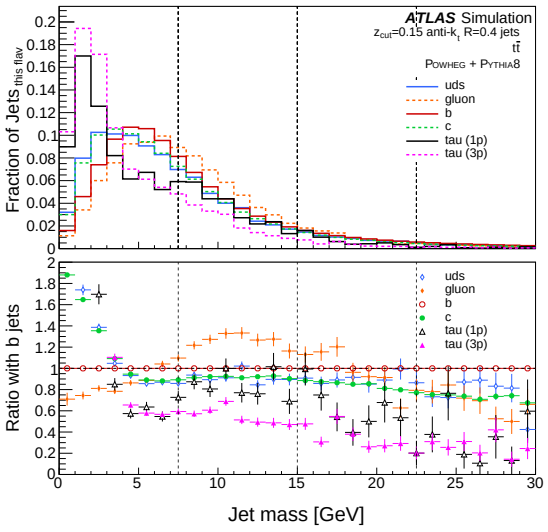


Figure D.5: Number of Clusters for tagged and untagged *b*-jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of tagged b jets.

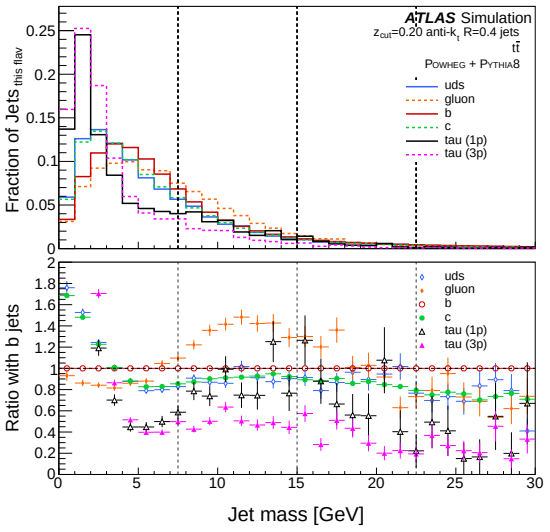


(a)

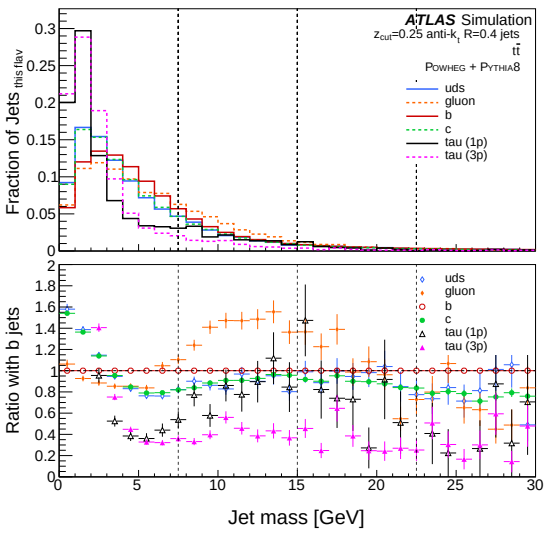
(b)



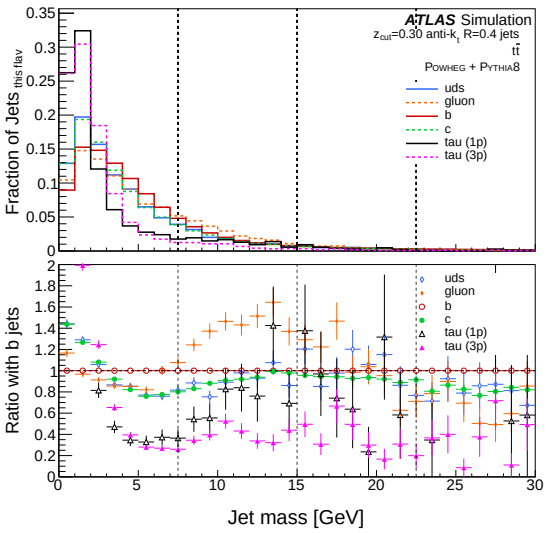
(c)



(d)



(e)



(f)

Figure D.6: Jet Mass for tagged and untagged *b*-jets, groomed with strength Z_{cut} = 0.05, (a), Z_{cut} = 0.10, (b), Z_{cut} = 0.15, (c), Z_{cut} = 0.20, (d), Z_{cut} = 0.25, (e), Z_{cut} = 0.30, (f), all normalised to the number of tagged b jets.

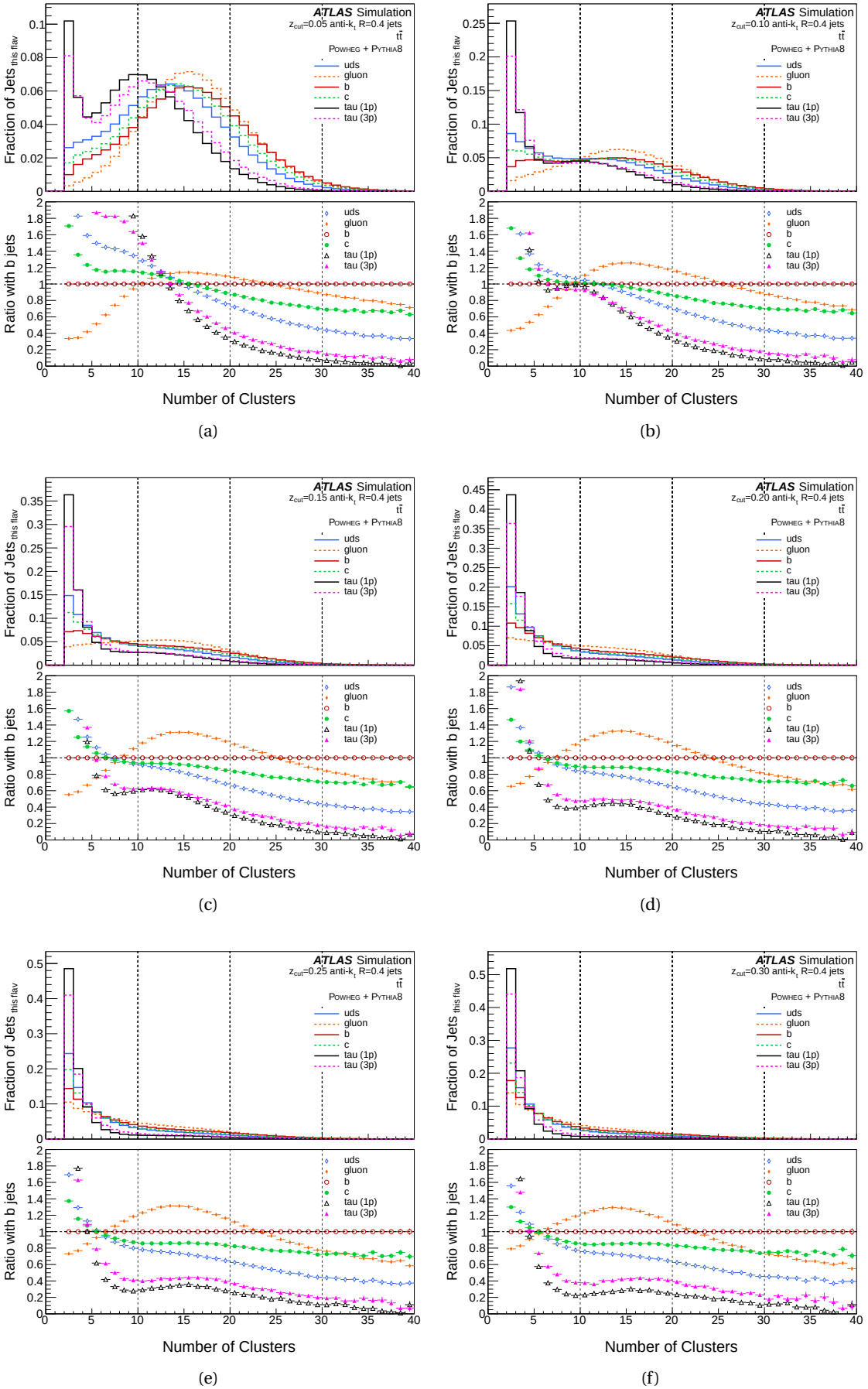
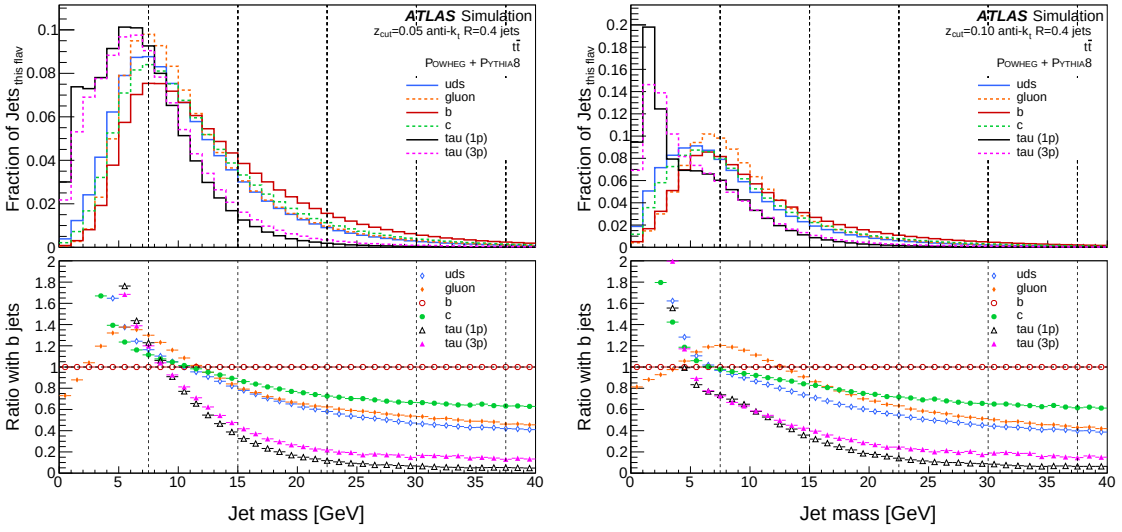
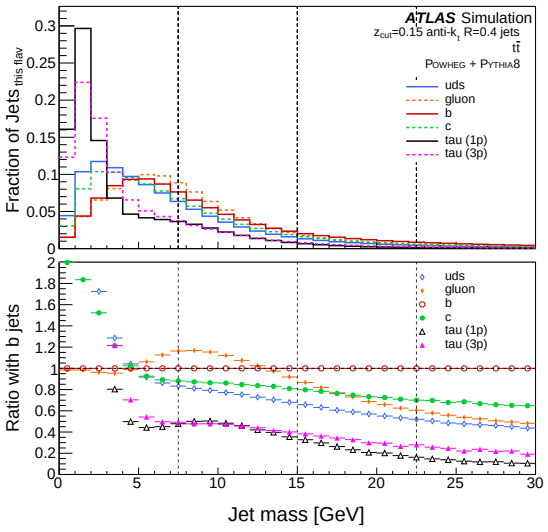


Figure D.7: Number of Clusters for tagged and untagged *b*-jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of tagged b jets.

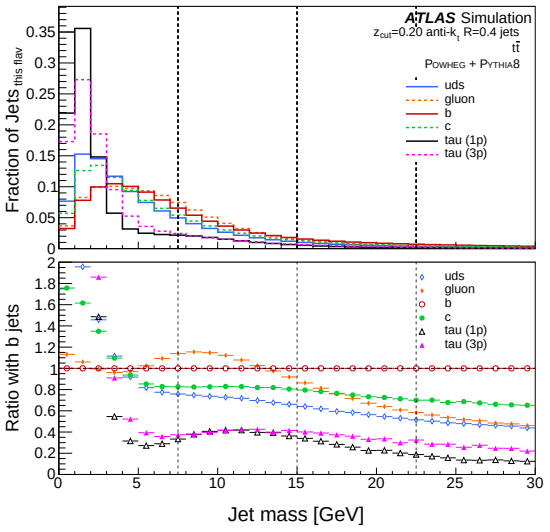


(a)

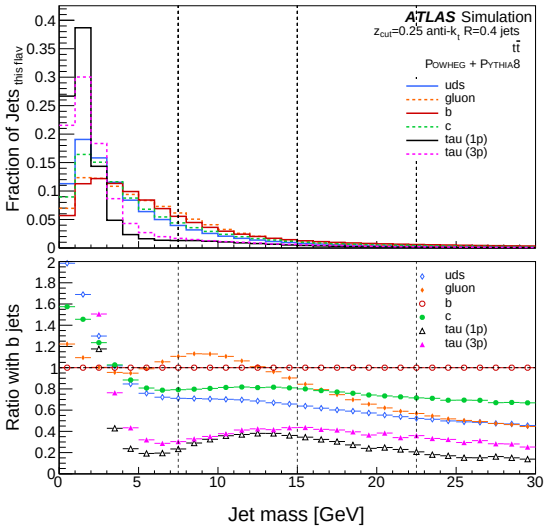
(b)



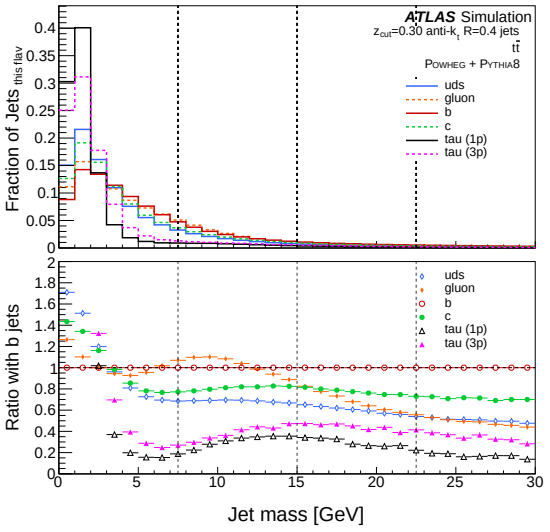
(c)



(d)

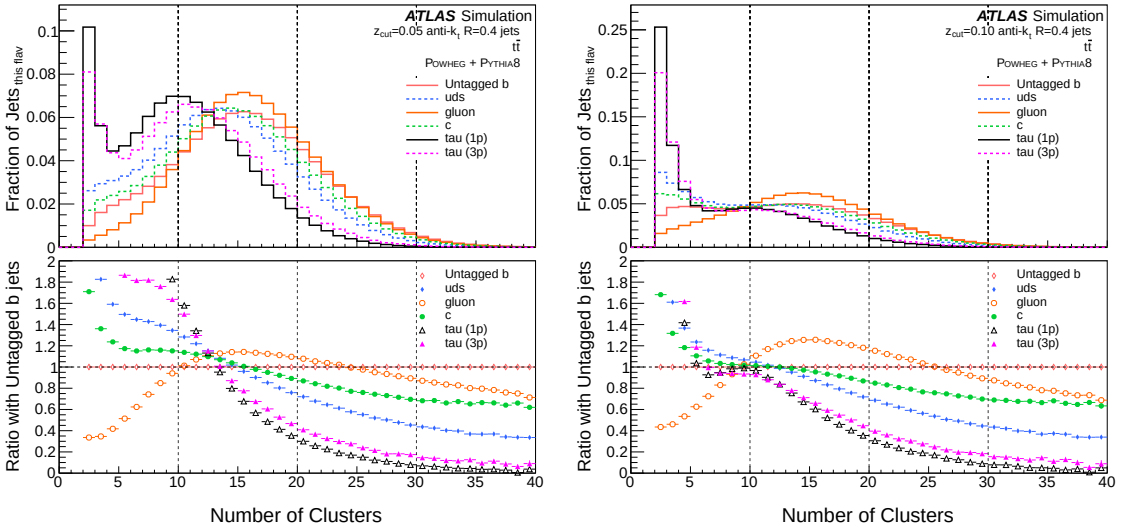


(e)



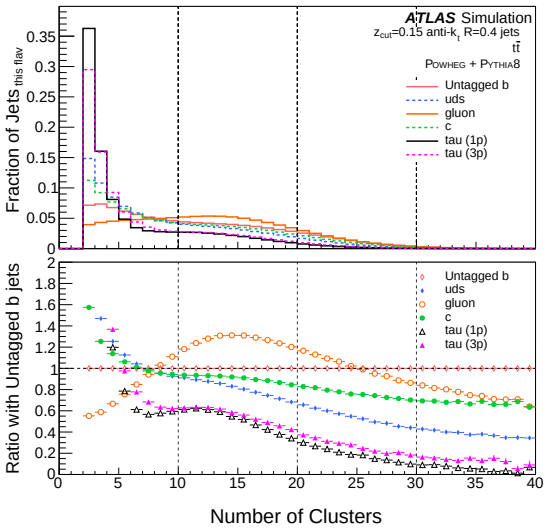
(f)

Figure D.8: Jet Mass for tagged and untagged *b*-jets, groomed with strength Z_{cut} = 0.05, (a), Z_{cut} = 0.10, (b), Z_{cut} = 0.15, (c), Z_{cut} = 0.20, (d), Z_{cut} = 0.25, (e), Z_{cut} = 0.30, (f), all normalised to the number of tagged b jets.

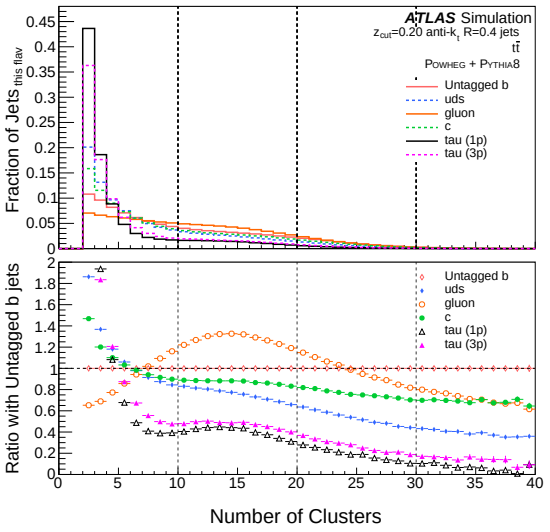


(a)

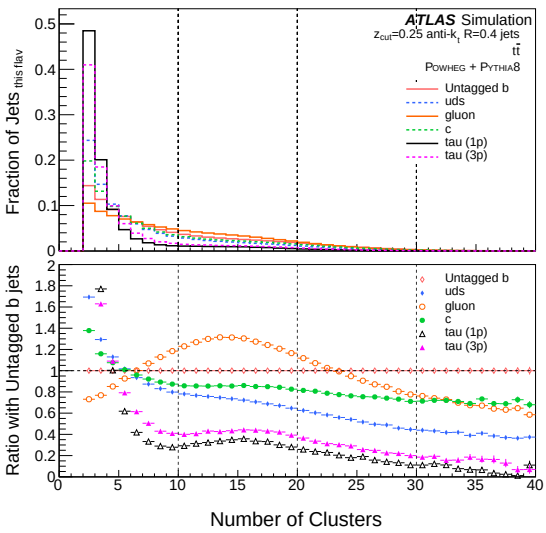
(b)



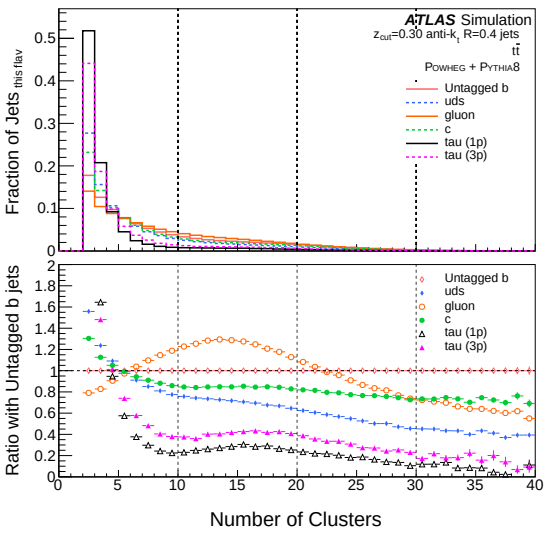
(c)



(d)

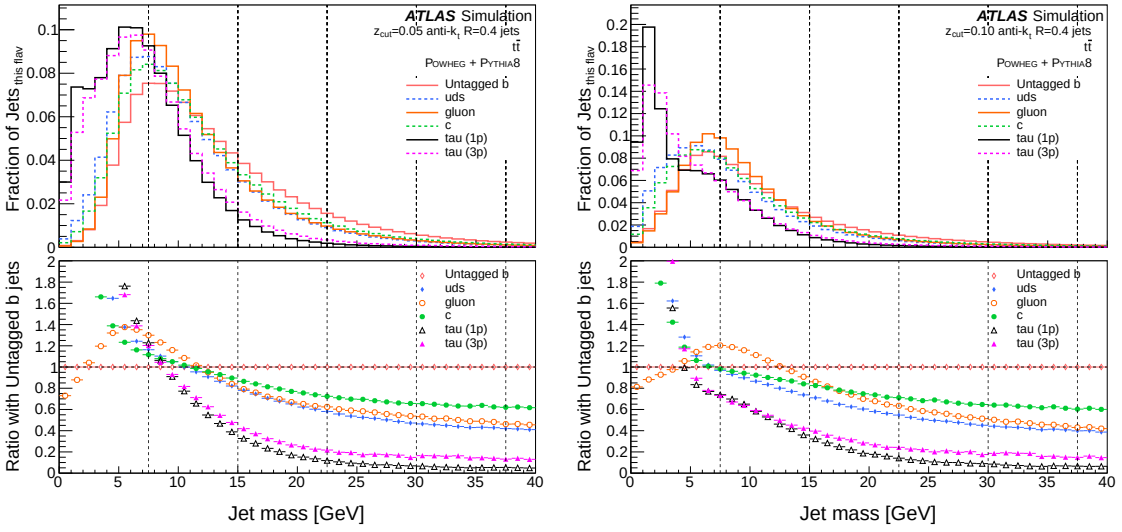


(e)



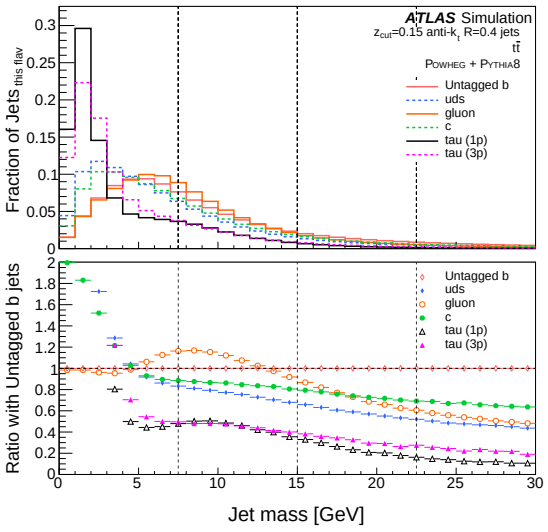
(f)

Figure D.9: Number of Clusters for tagged and untagged *b*-jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of tagged b jets.

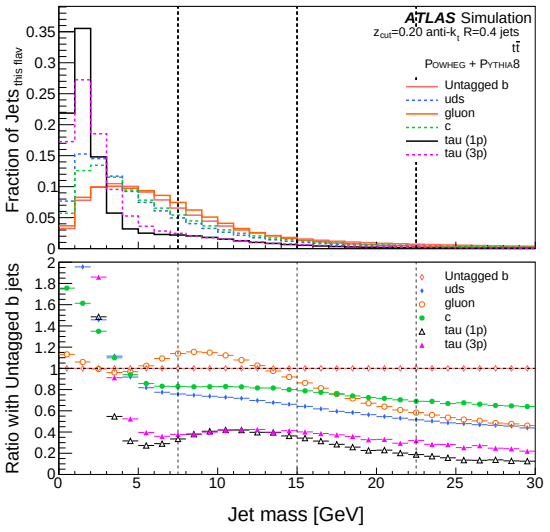


(a)

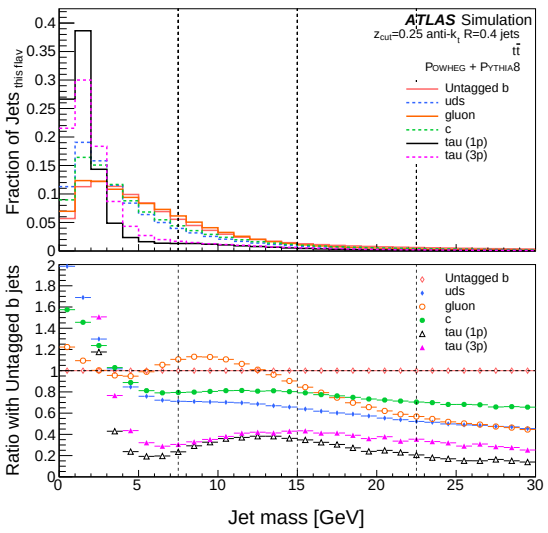
(b)



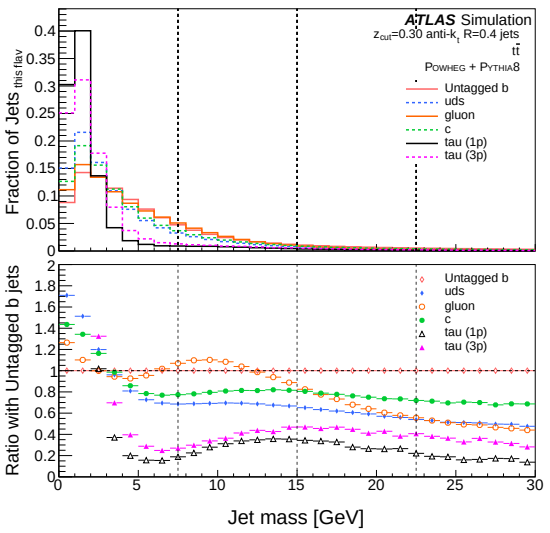
(c)



(d)

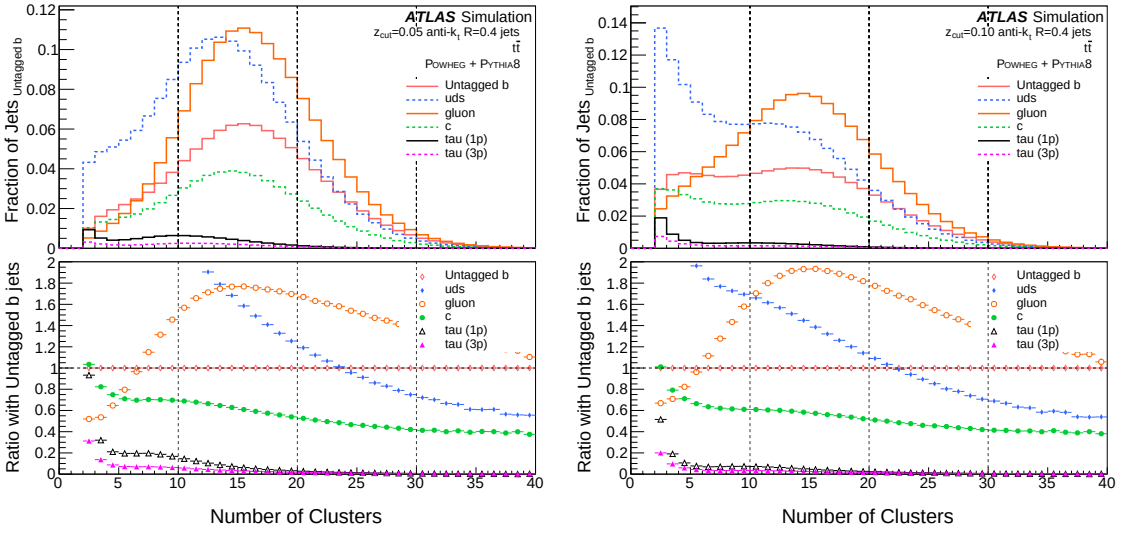


(e)



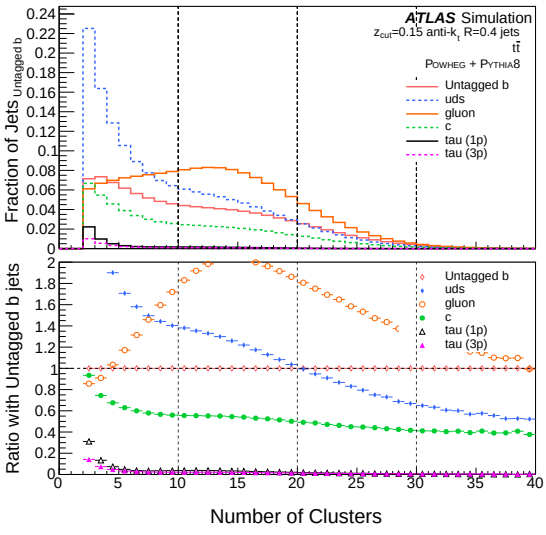
(f)

Figure D.10: Jet Mass for tagged and untagged b-jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of tagged b jets.

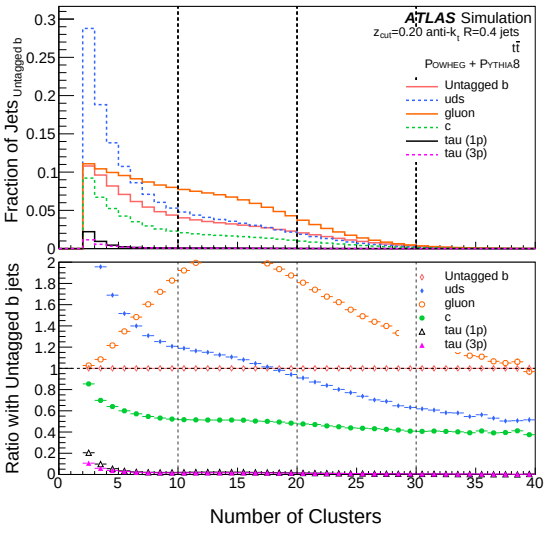


(a)

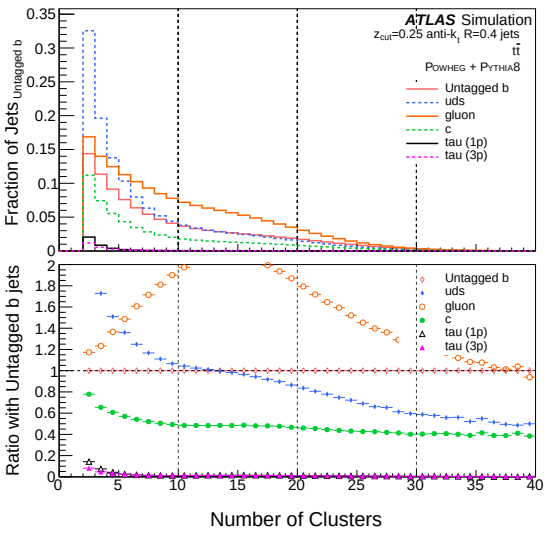
(b)



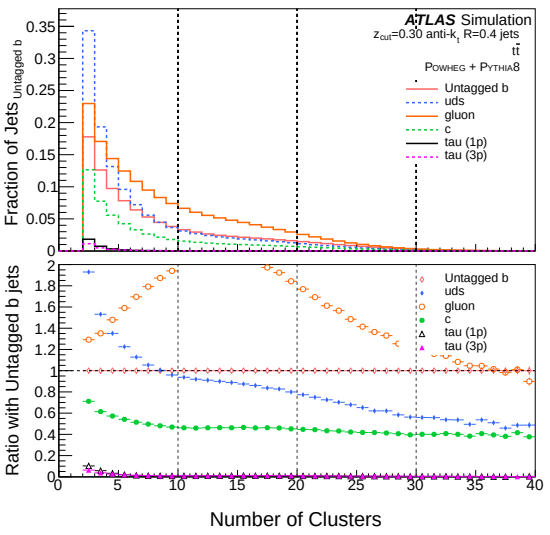
(c)



(d)

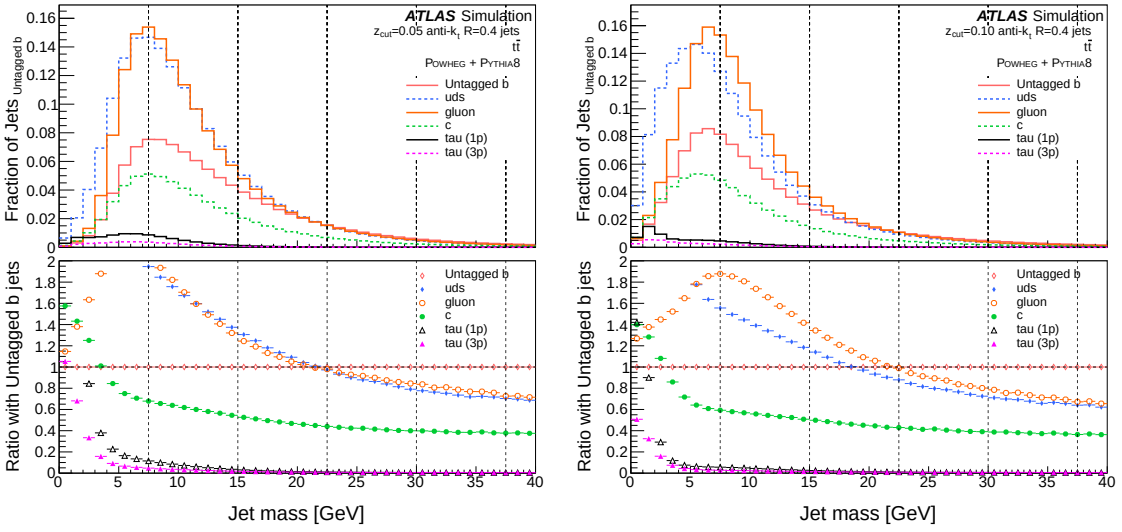


(e)



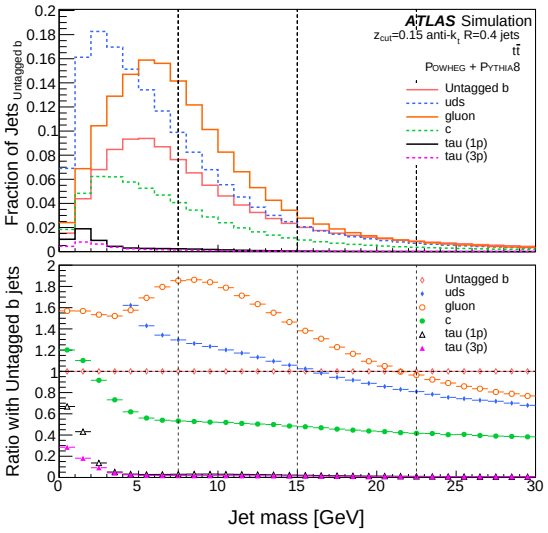
(f)

Figure D.11: Number of Clusters for tagged and untagged *b*-jets, groomed with strength $Z_{cut} = 0.05$, (a), $Z_{cut} = 0.10$, (b), $Z_{cut} = 0.15$, (c), $Z_{cut} = 0.20$, (d), $Z_{cut} = 0.25$, (e), $Z_{cut} = 0.30$, (f), all normalised to the number of tagged b jets.

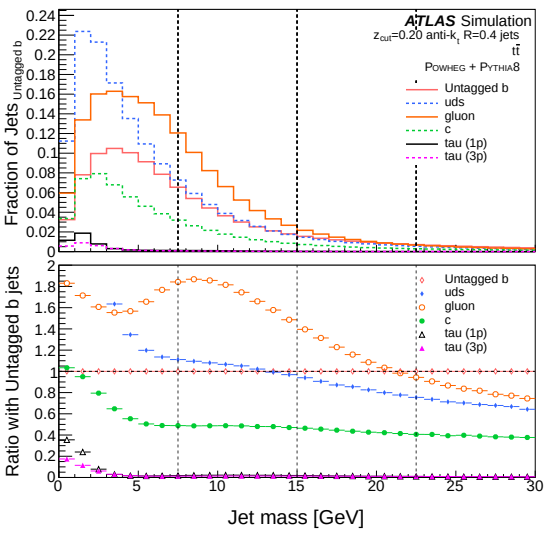


(a)

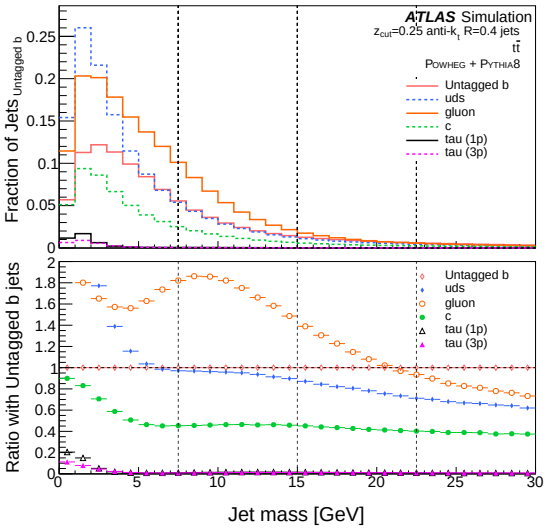
(b)



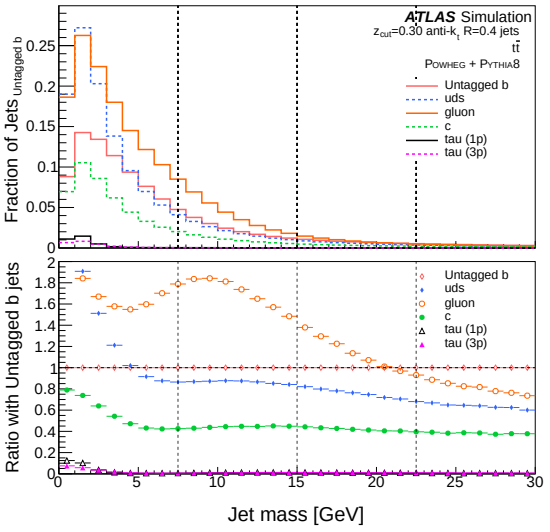
(c)



(d)



(e)



(f)

Figure D.12: Jet Mass for tagged and untagged *b*-jets, groomed with strength Z $_{cut}$ = 0.05, (a), Z $_{cut}$ = 0.10, (b), Z $_{cut}$ = 0.15, (c), Z $_{cut}$ = 0.20, (d), Z $_{cut}$ = 0.25, (e), Z $_{cut}$ = 0.30, (f), all normalised to the number of tagged b jets.

E

SOFT DROP GROOMING ON FLAVOUR-TAGGED SMALL JETS() FROM $t\bar{t}H$ AND $t\bar{t}$ EVENTS.

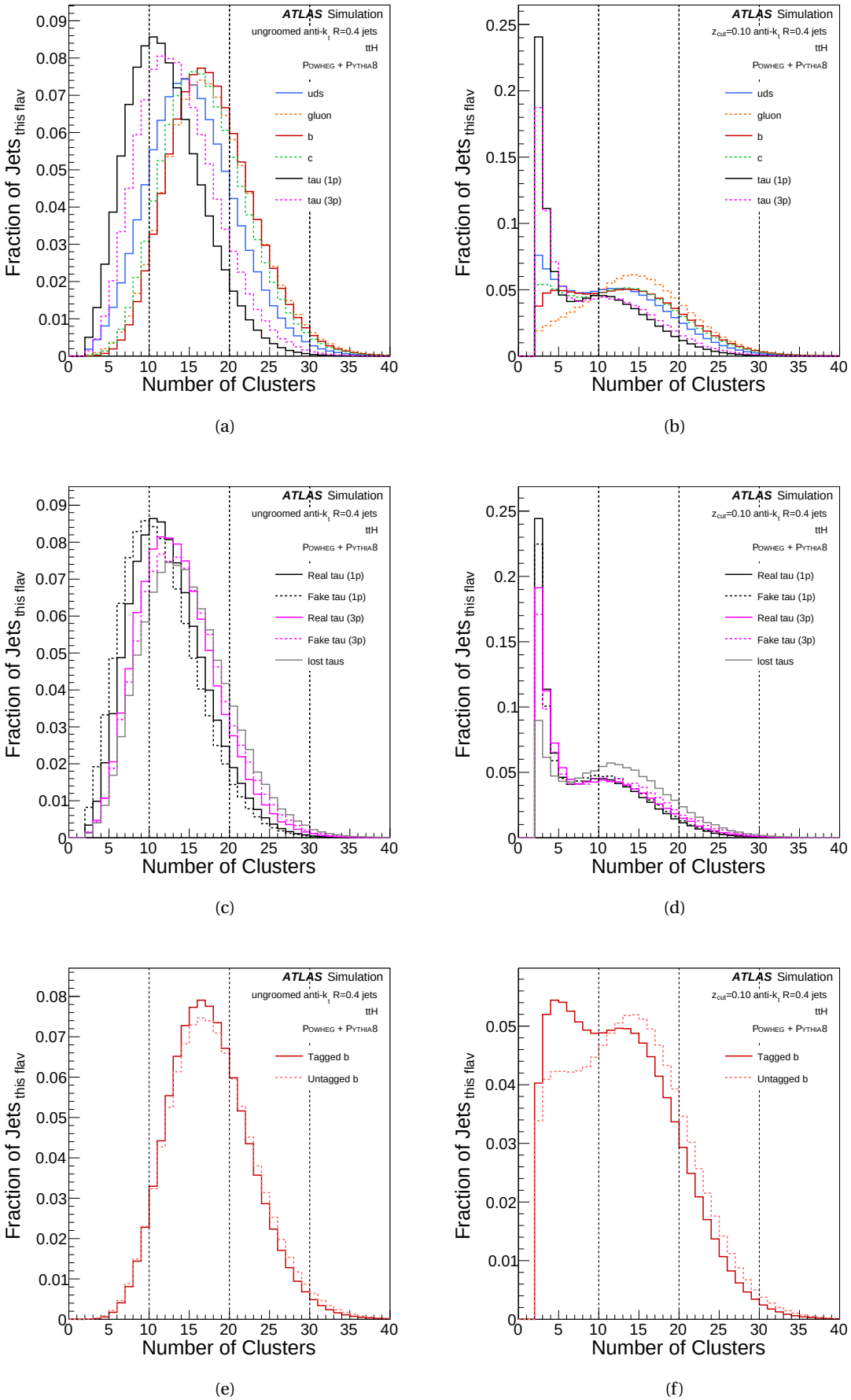


Figure E.1: NCLUS for $t\bar{t}H$ jets of different flavours

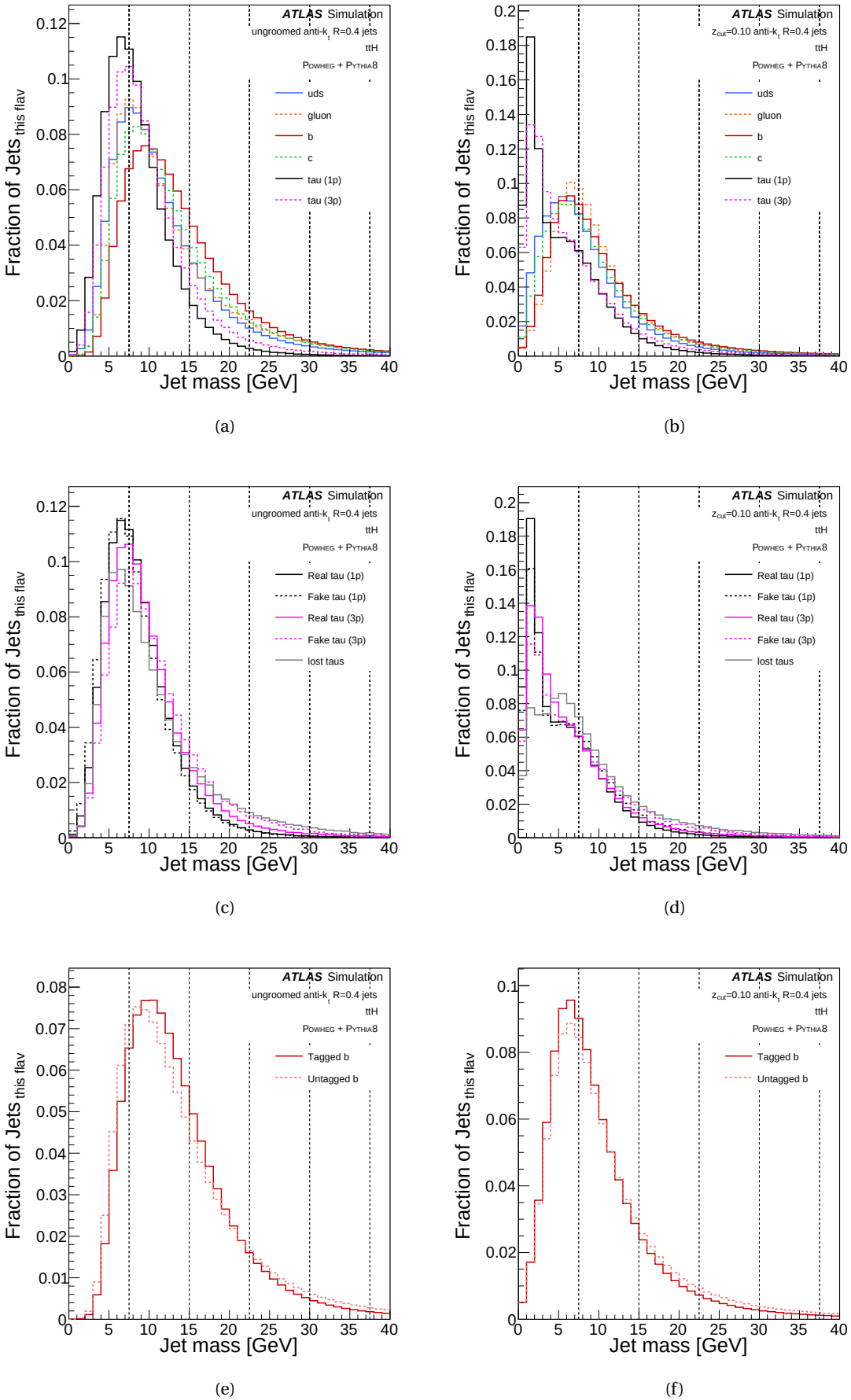
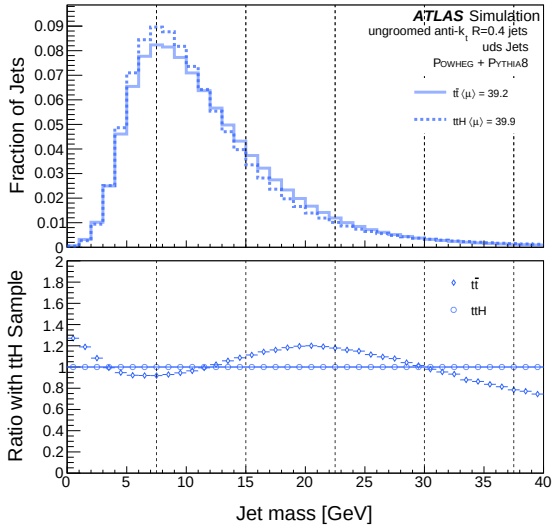
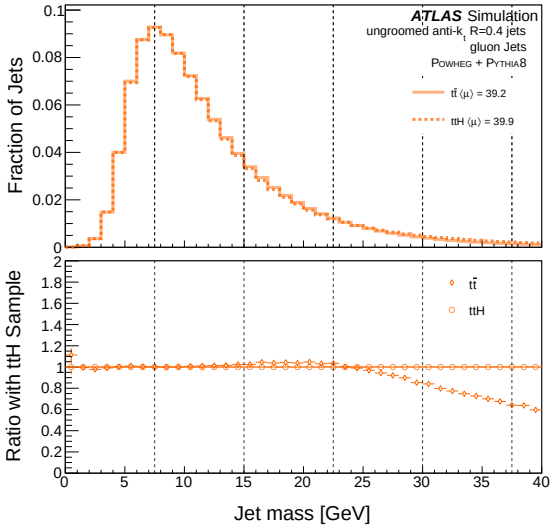


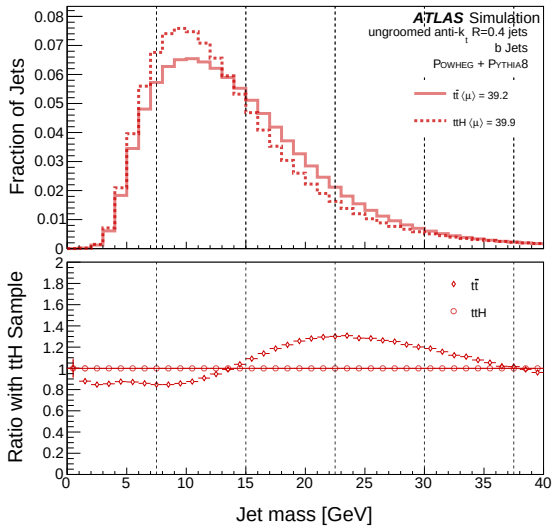
Figure E.2: Jet mass for $t\bar{t}H$ jets of different flavours



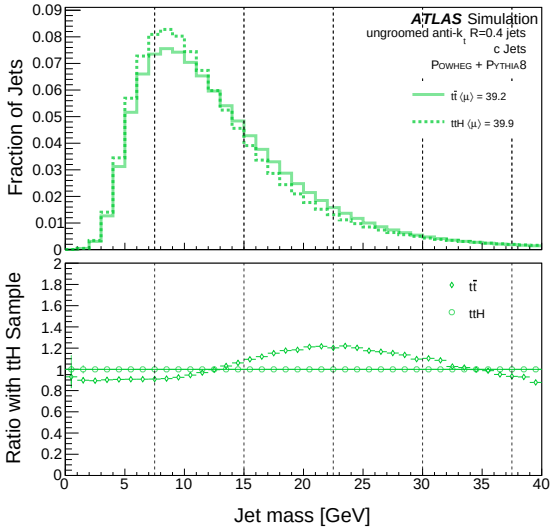
(a)



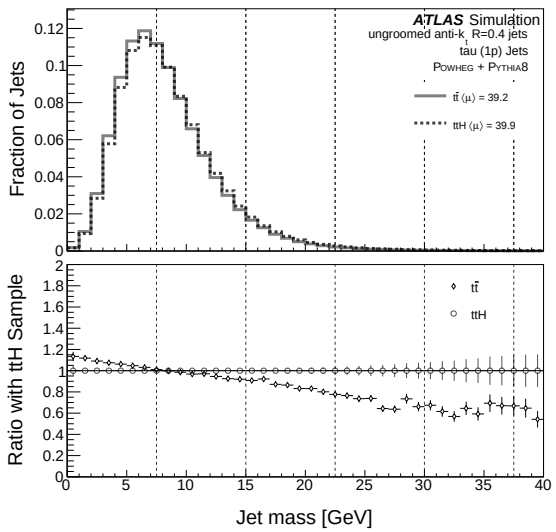
(b)



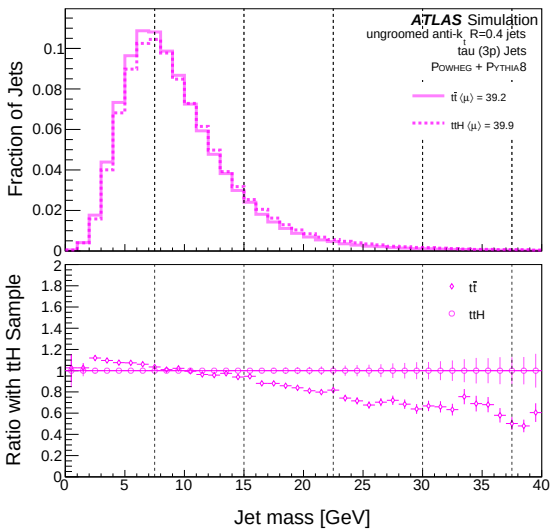
(c)



(d)

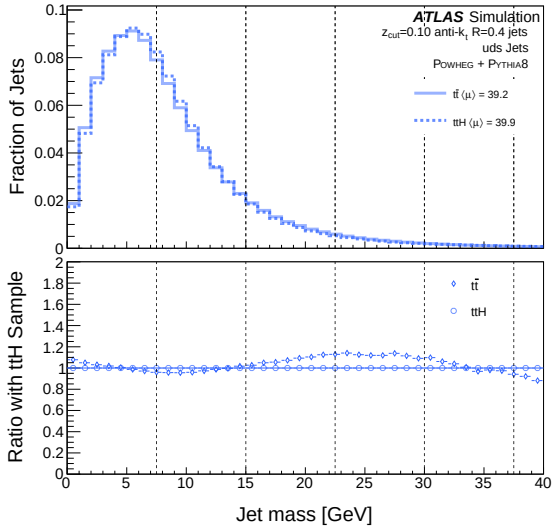


(e)

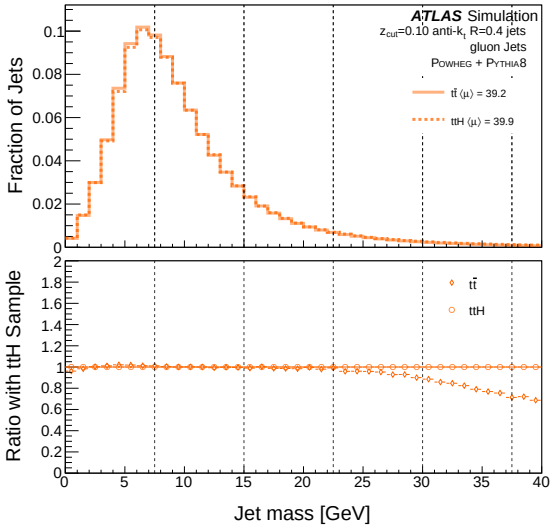


(f)

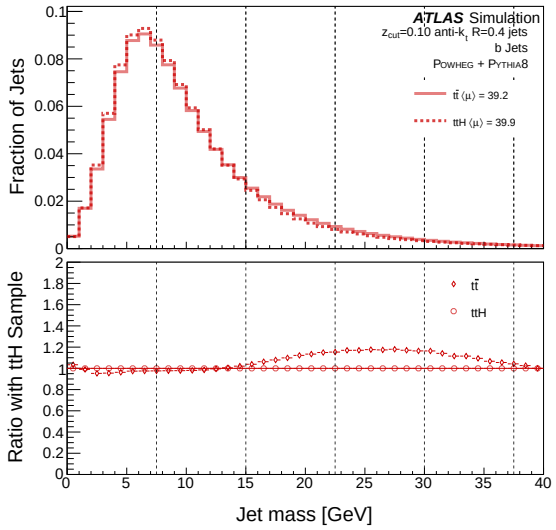
Figure E.3: Jet mass for ungroomed $t\bar{t}H$ and $t\bar{t}$ jets of different flavours



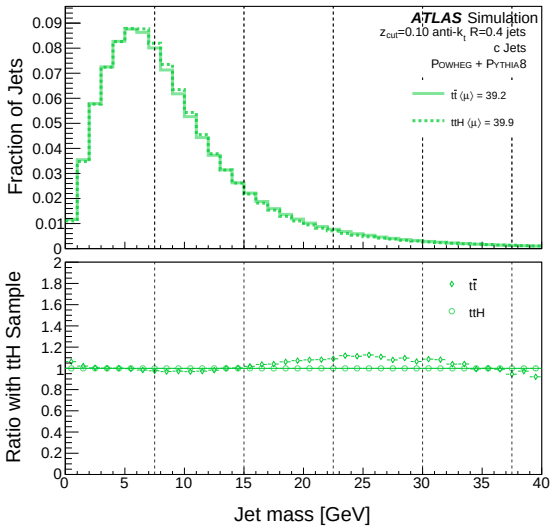
(a)



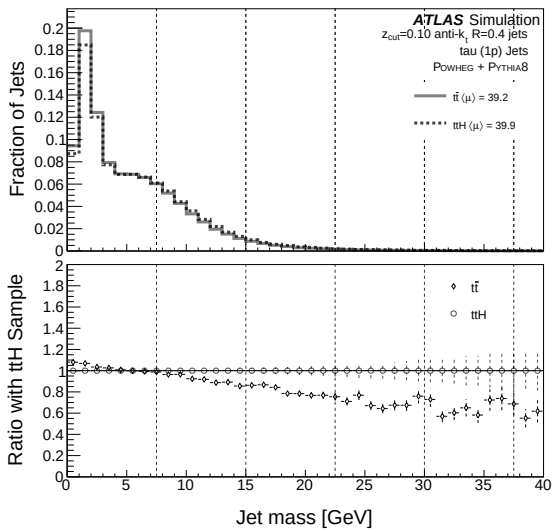
(b)



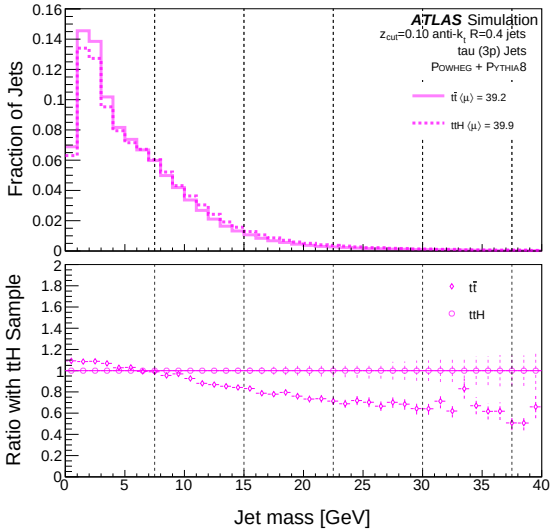
(c)



(d)

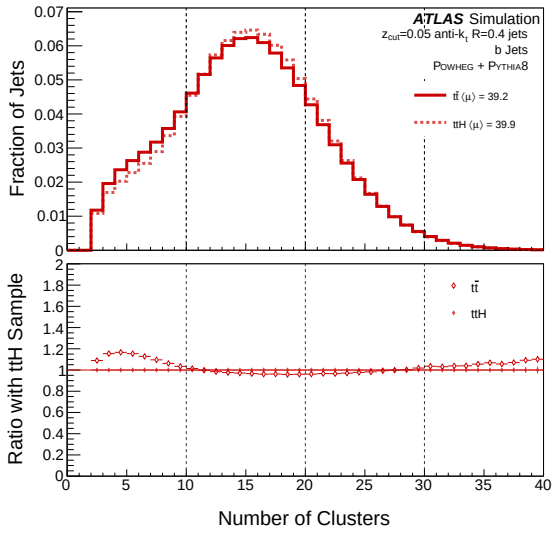


(e)

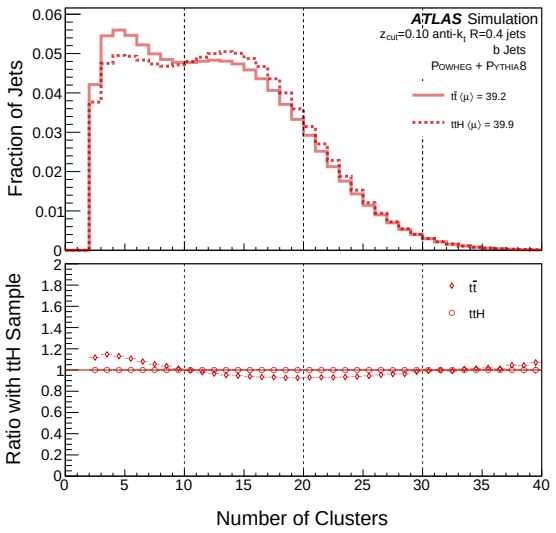


(f)

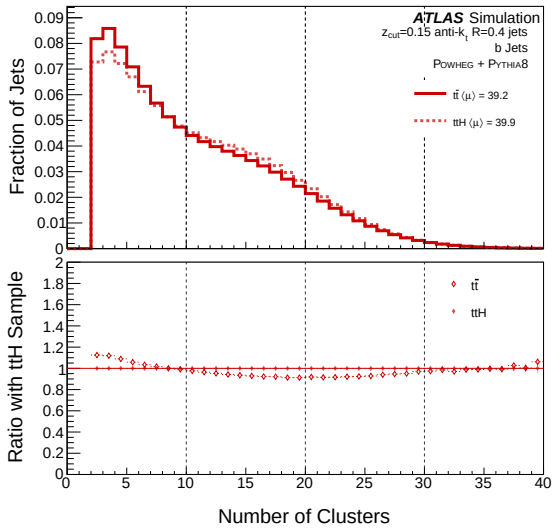
Figure E.4: Jet mass for $t\bar{t}H$ and $t\bar{t}$ jets of different flavours with soft drop grooming of strength $Z_{cut} = 0.10$ applied



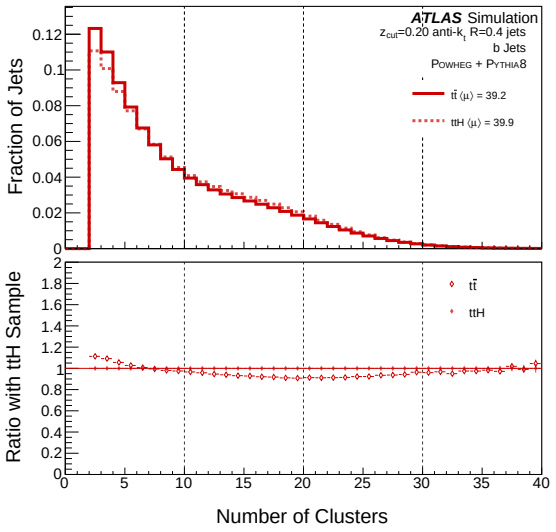
(a)



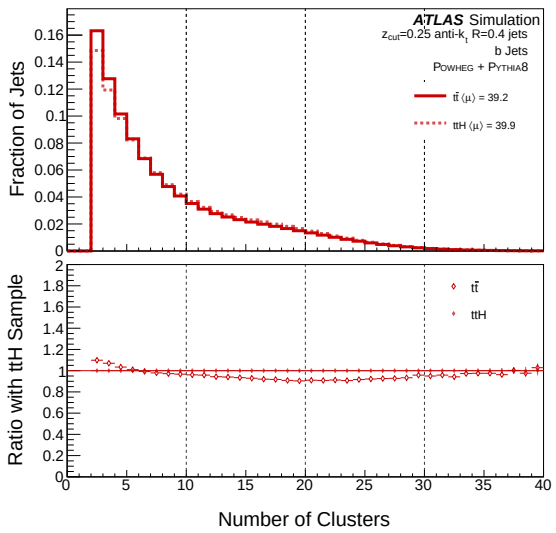
(b)



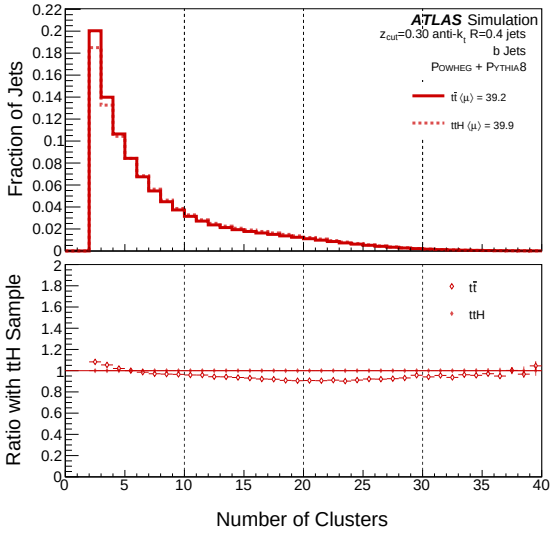
(c)



(d)

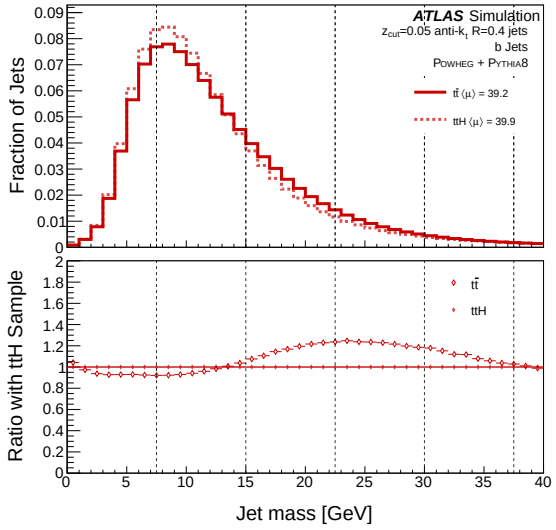


(e)

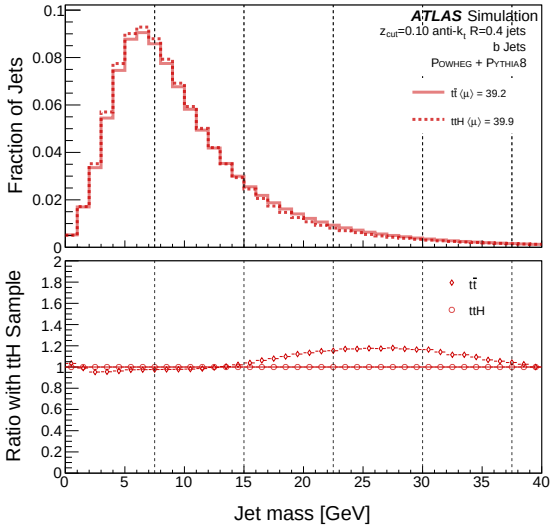


(f)

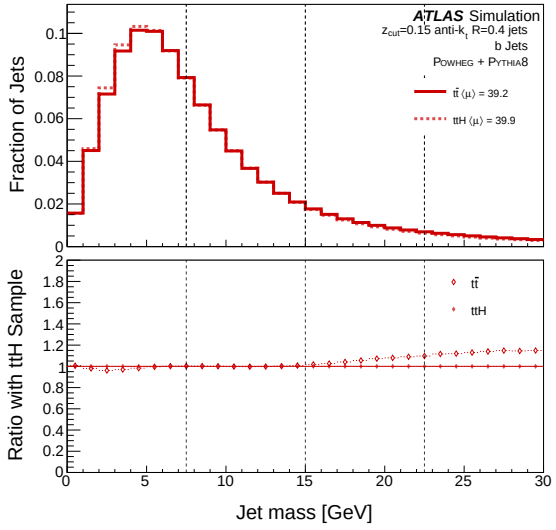
Figure E.5: Sample comparison of Number of Clusters for groomed b-jets in $t\bar{t}H$ and $t\bar{t}$



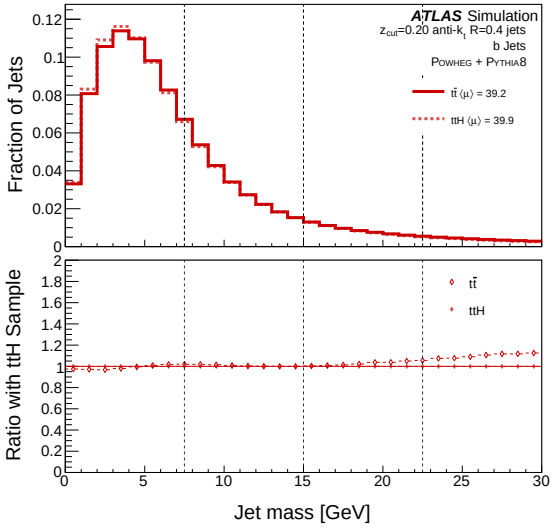
(a)



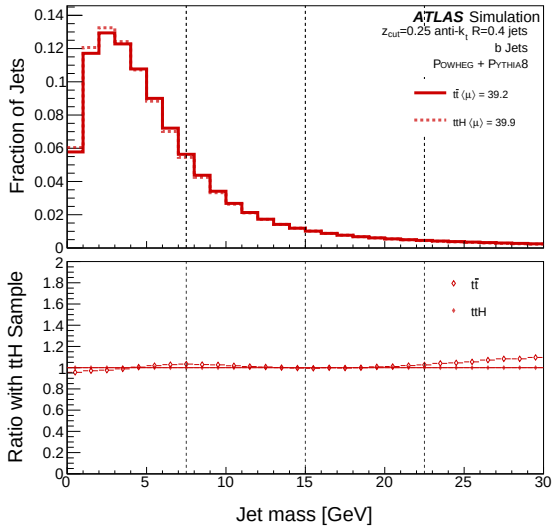
(b)



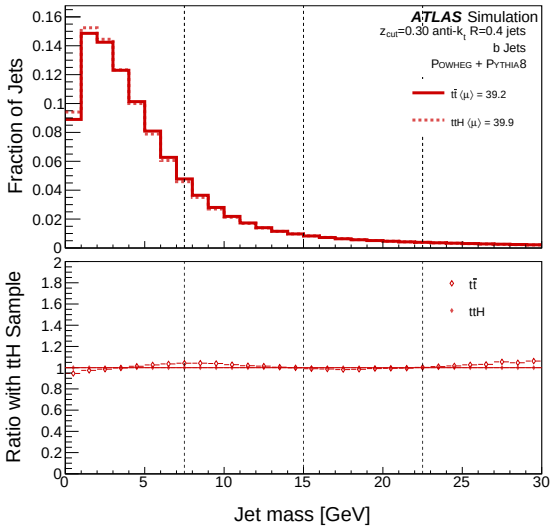
(c)



(d)

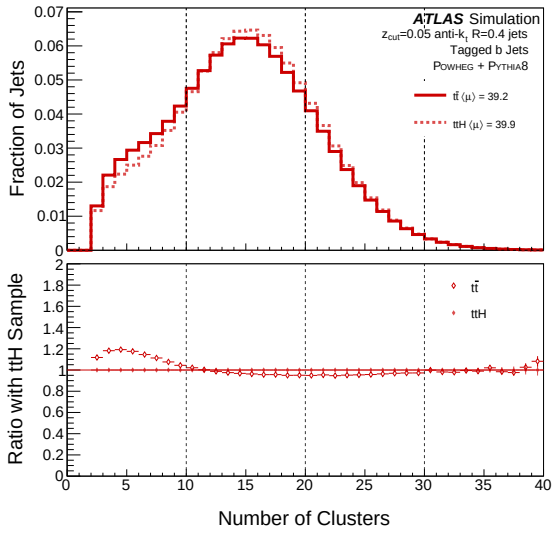


(e)

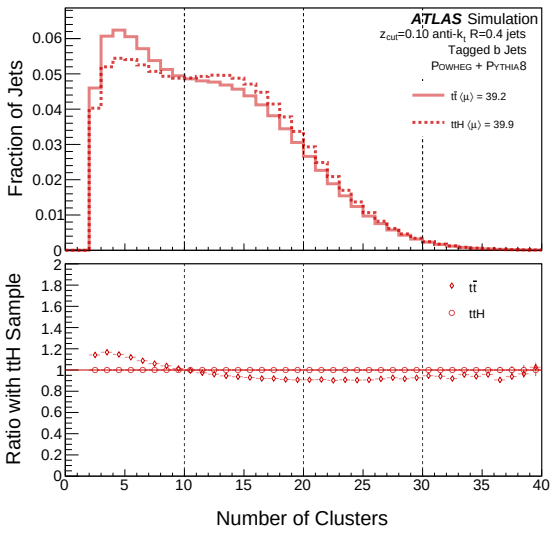


(f)

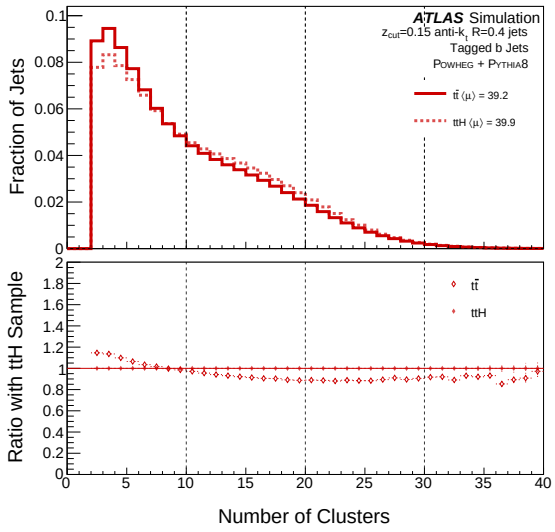
Figure E.6: Sample comparison of jet mass for groomed b-jets in $t\bar{t}H$ and $t\bar{t}$



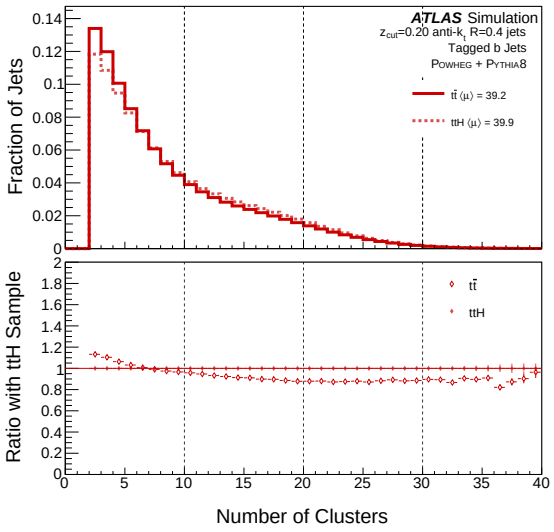
(a)



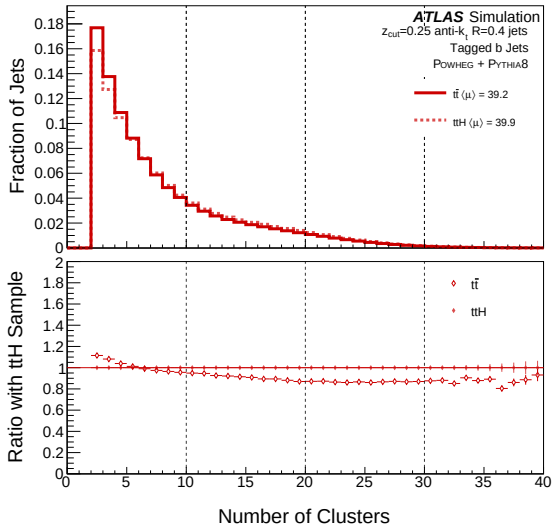
(b)



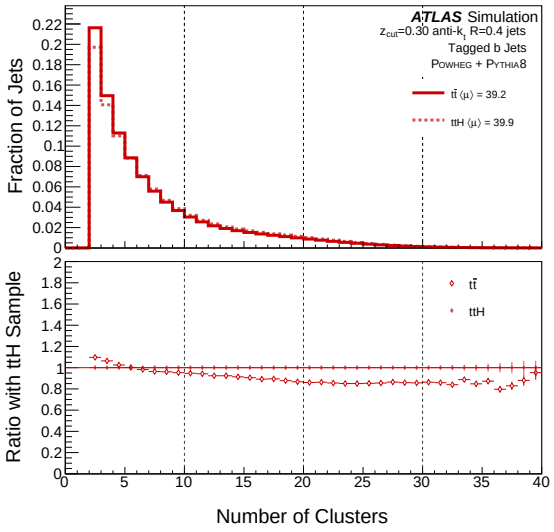
(c)



(d)

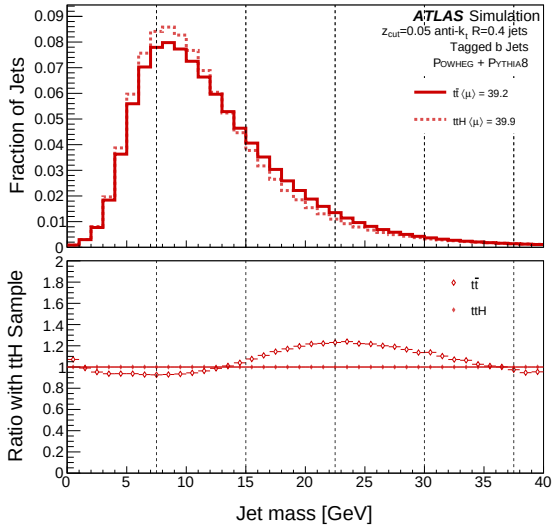


(e)

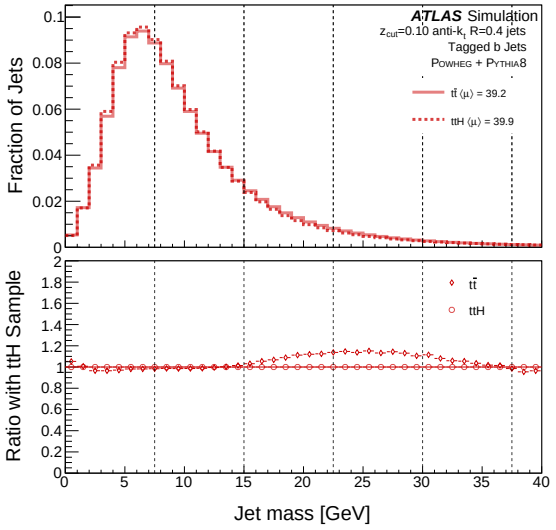


(f)

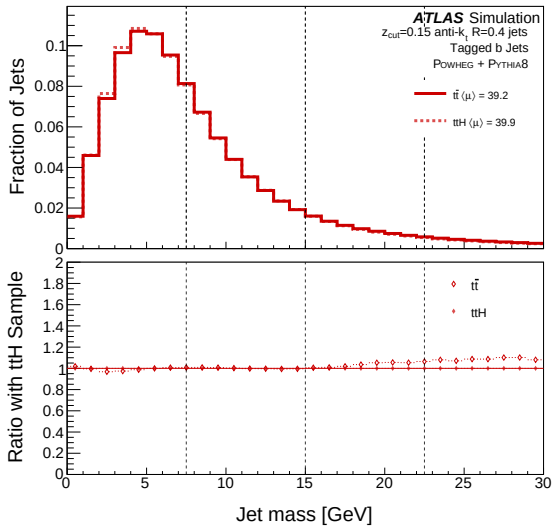
Figure E.7: Sample comparison of Number of Clusters for groomed b-jets in $t\bar{t}H$ and $t\bar{t}$



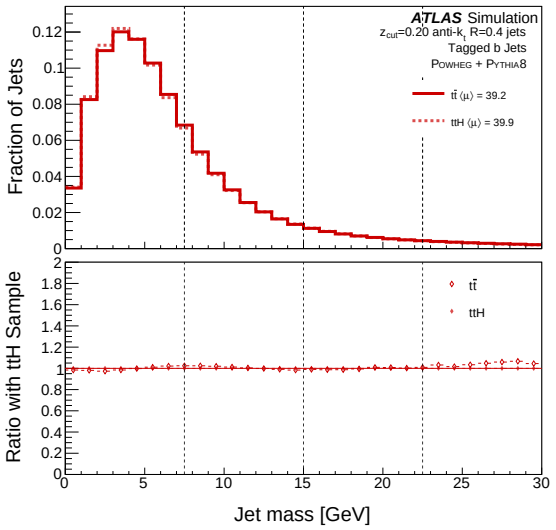
(a)



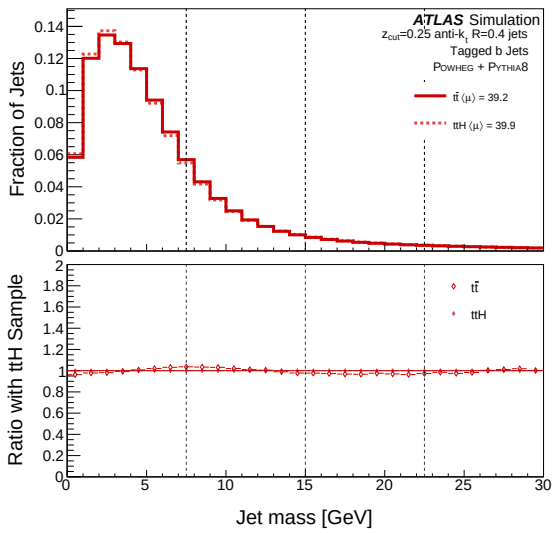
(b)



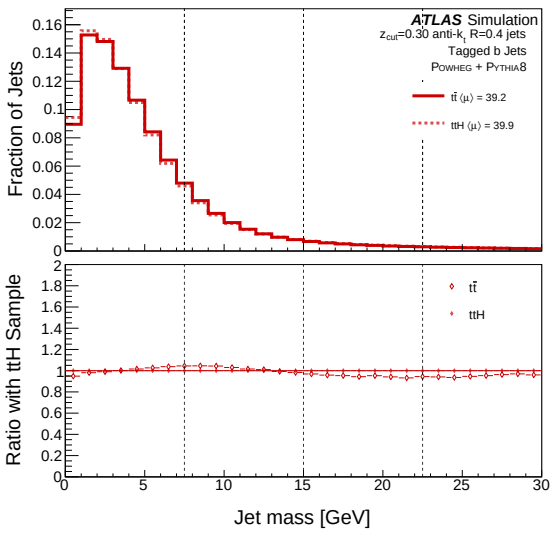
(c)



(d)



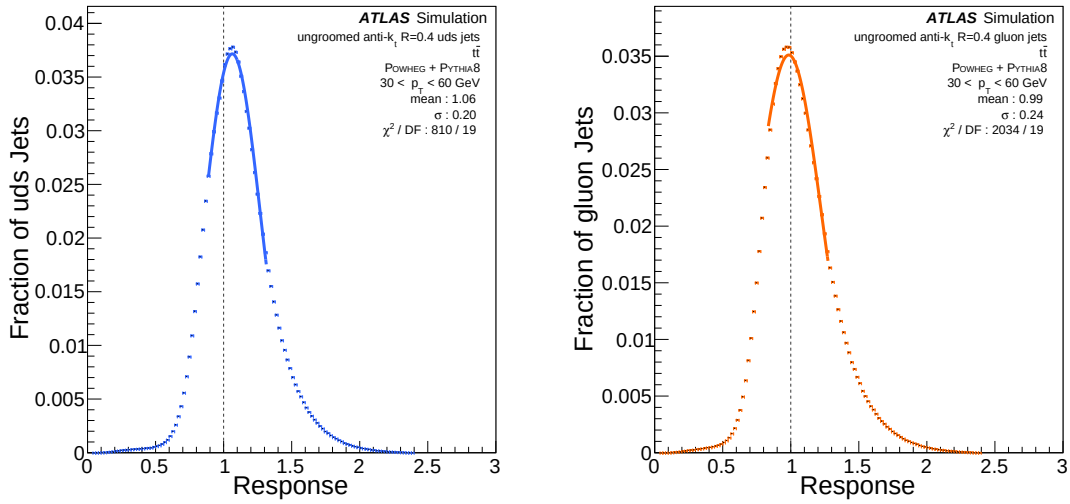
(e)



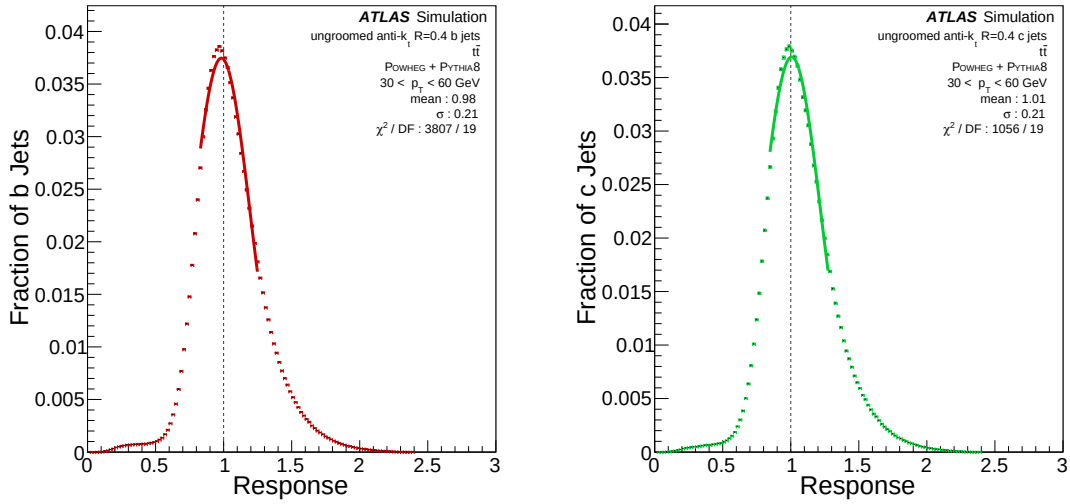
(f)

Figure E.8: Sample comparison of jet mass for groomed b-jets in $t\bar{t}H$ and $t\bar{t}$

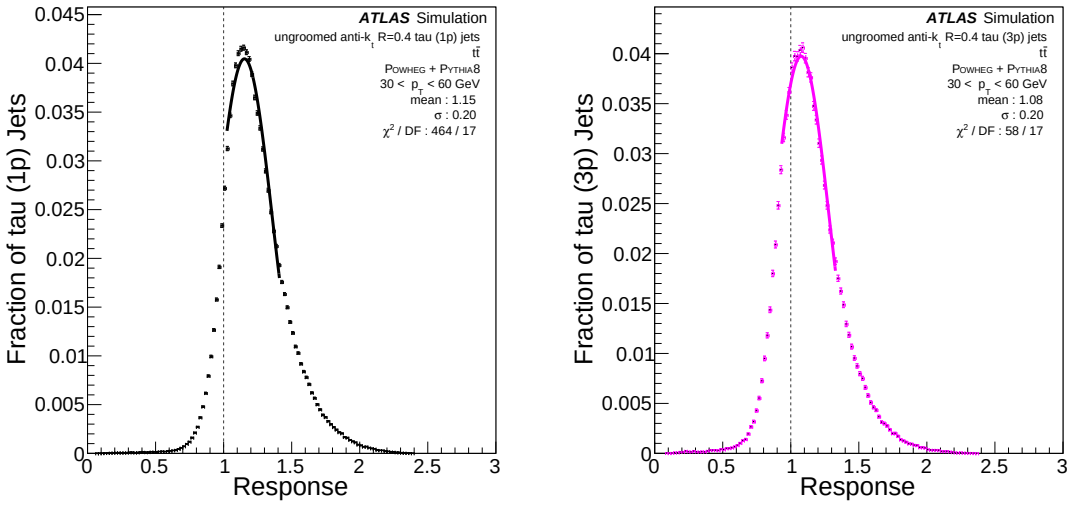
JES AND JER



(a) (b)



(c) (d)



(e) (f)

Figure F.1: Jet Response of ungroomed uds-, (a), gluon-, (b), b-, (c), c-, (d), 1-pronged tau -, (e) and 3-pronged tau -jets, (f)

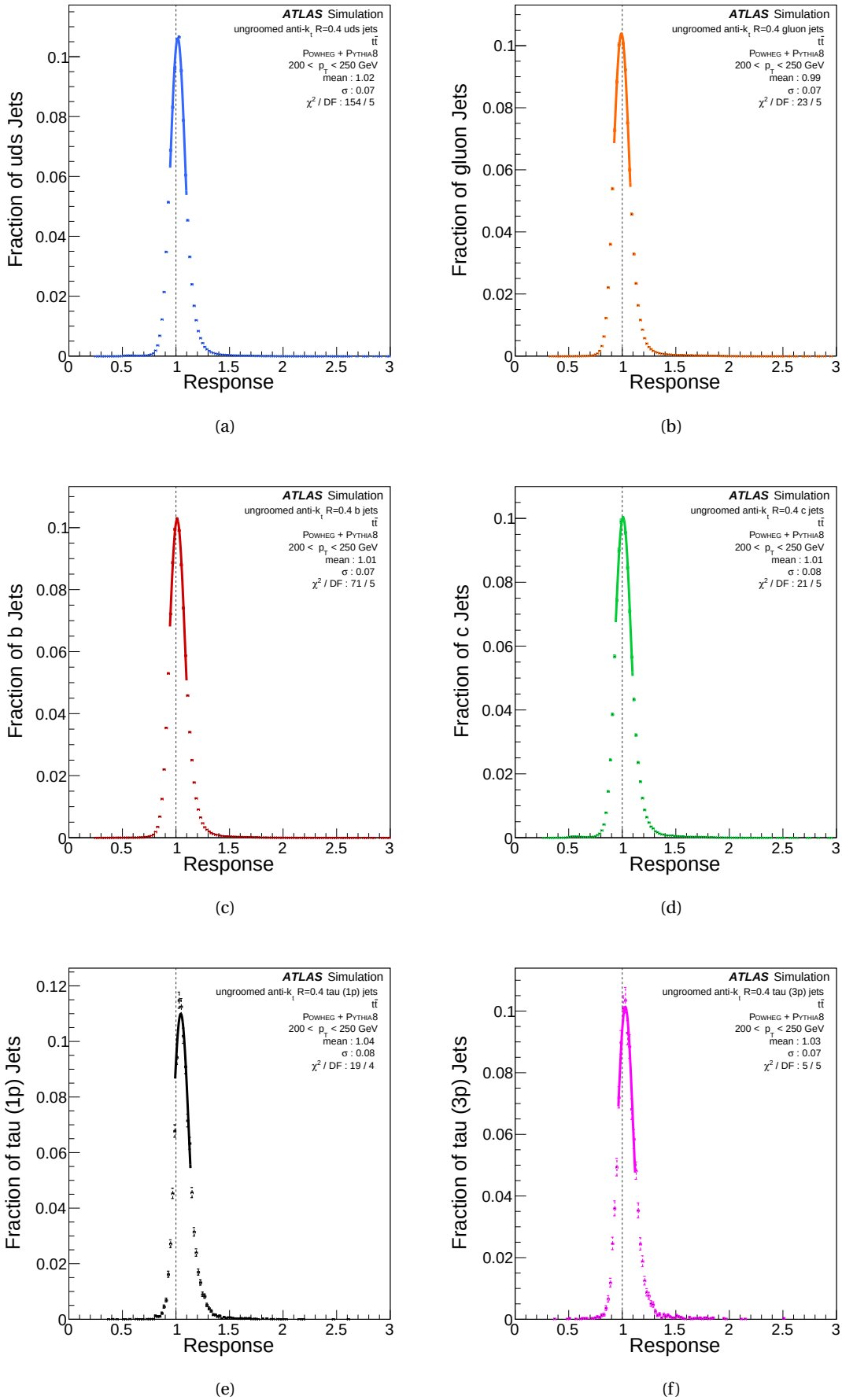


Figure F.2: Jet Response of ungroomed uds-, (a), gluon-, (b), b-, (c), c-, (d), 1-pronged tau -, (e) and 3-pronged tau -jets, (f)

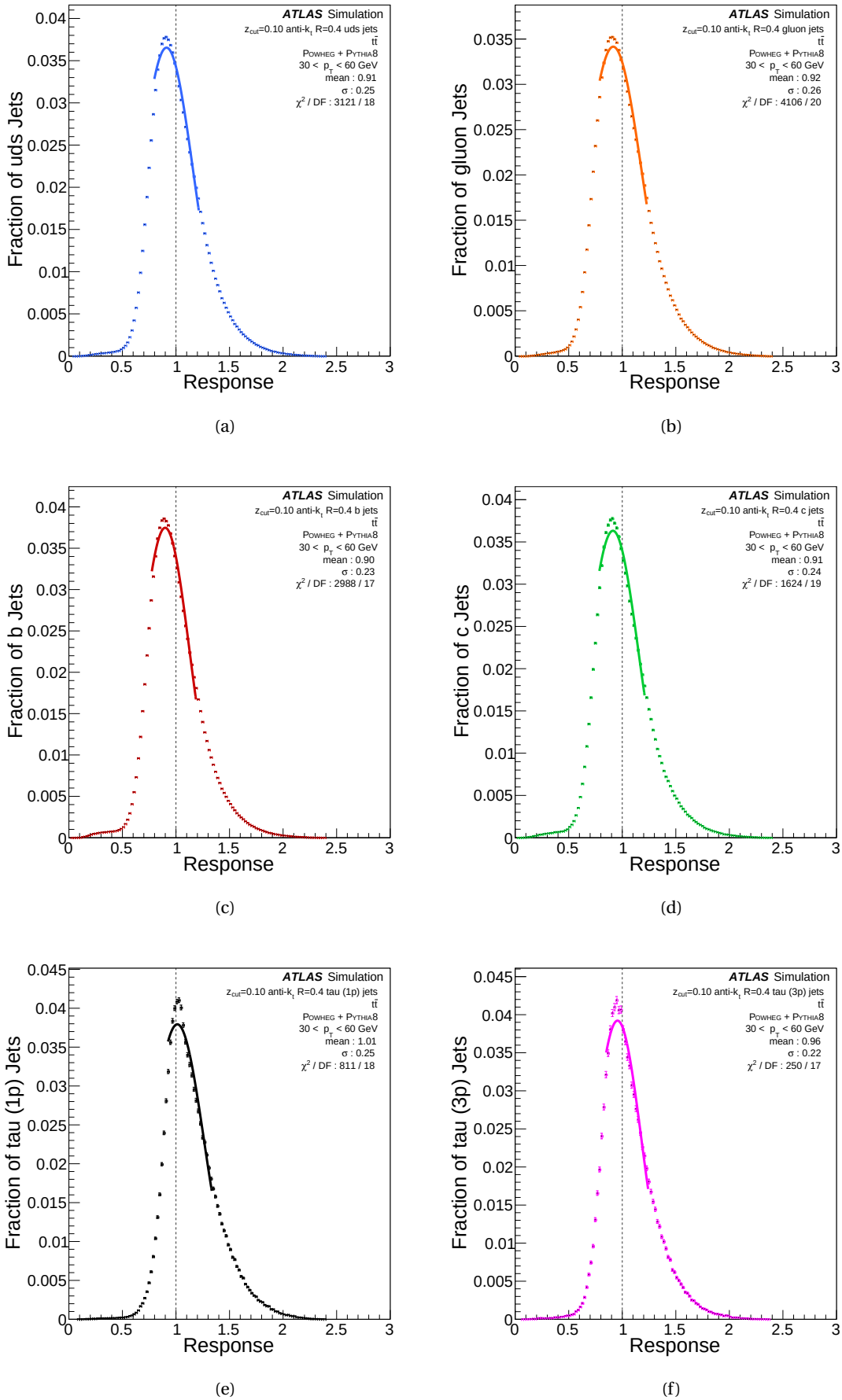


Figure F.3: Jet Response of SDb0z1 uds-, (a), gluon-, (b), b-, (c), c-, (d), 1-pronged tau -, (e) and 3-pronged tau -jets, (f)

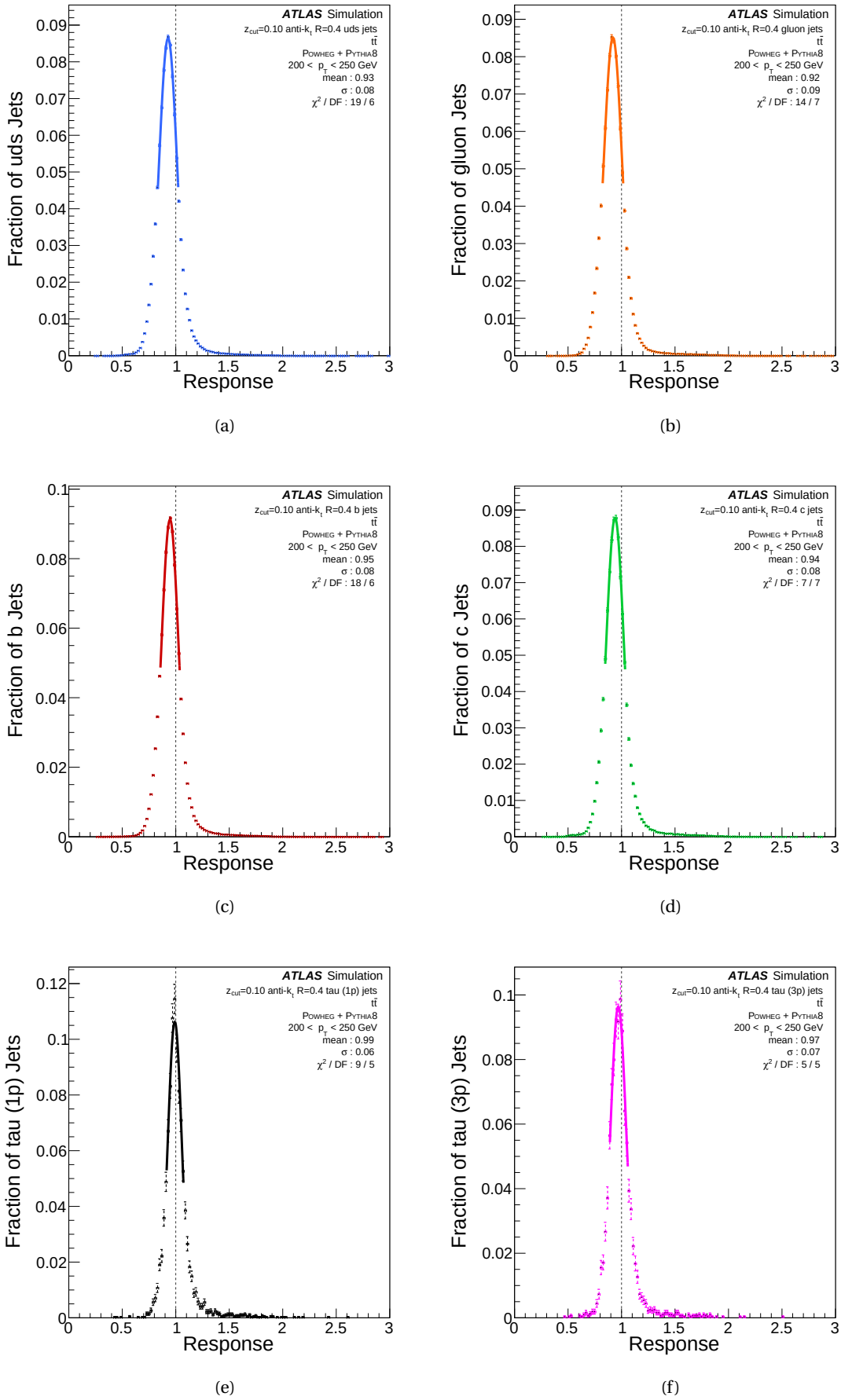
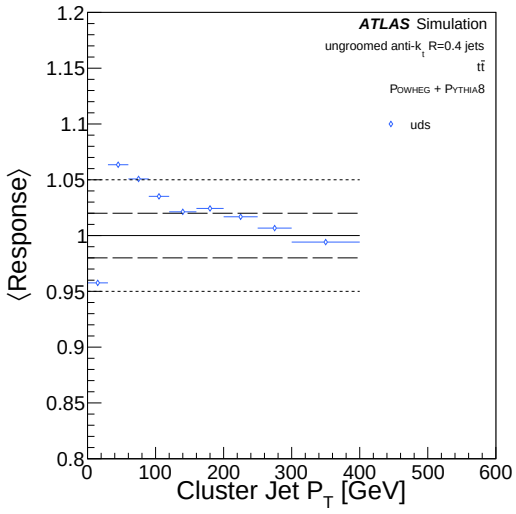
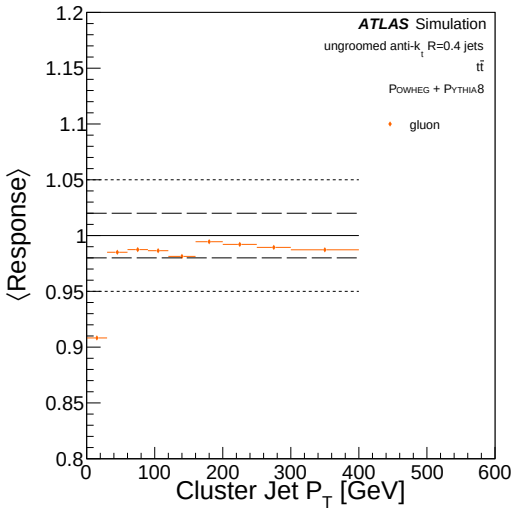


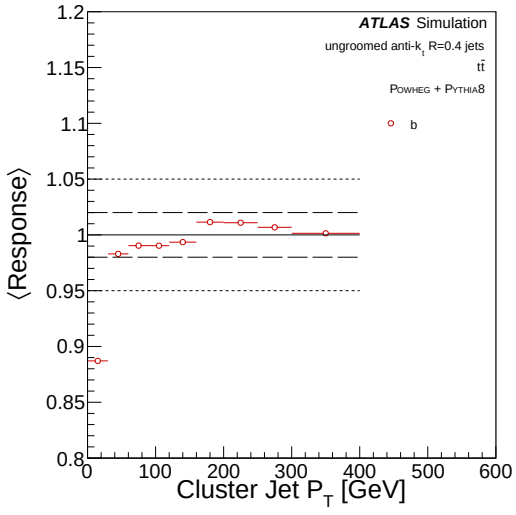
Figure F.4: Jet Response of groomed uds-, (a), gluon-, (b), b-, (c), c-, (d), 1-pronged tau -, (e) and 3-pronged tau -jets, (f)



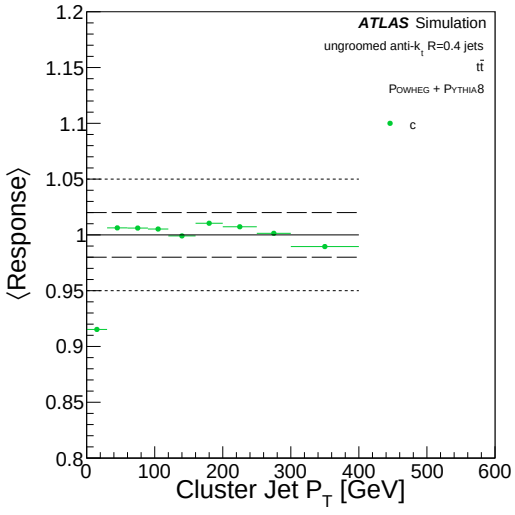
(a)



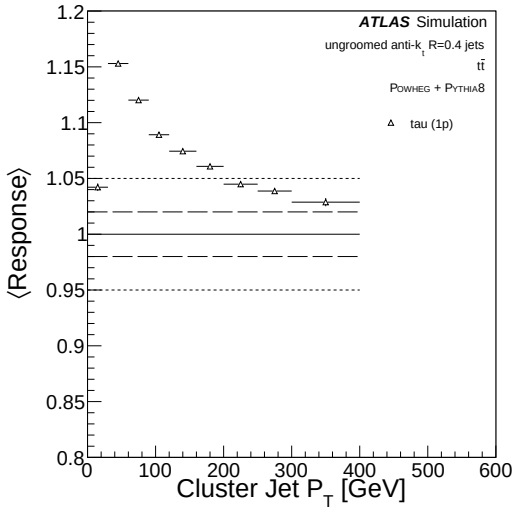
(b)



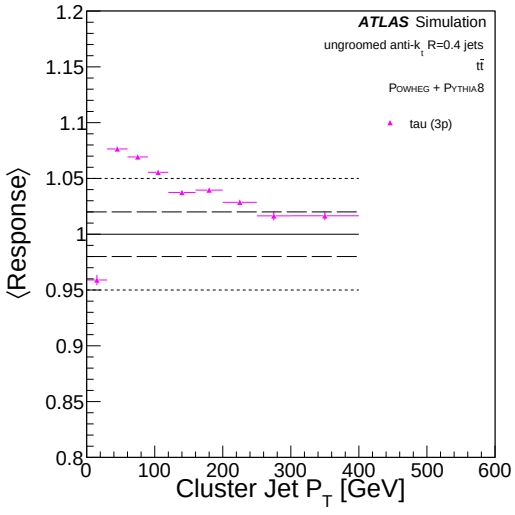
(c)



(d)

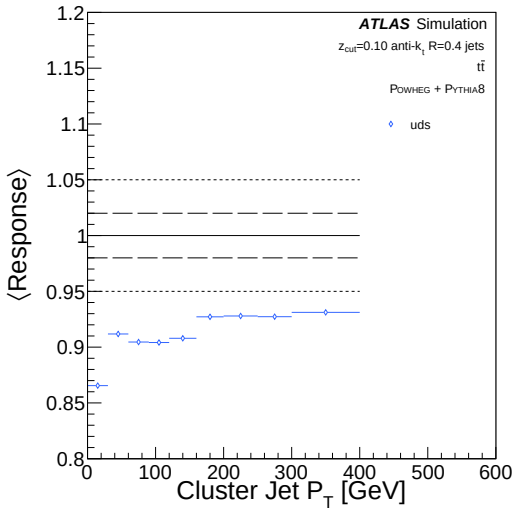


(e)

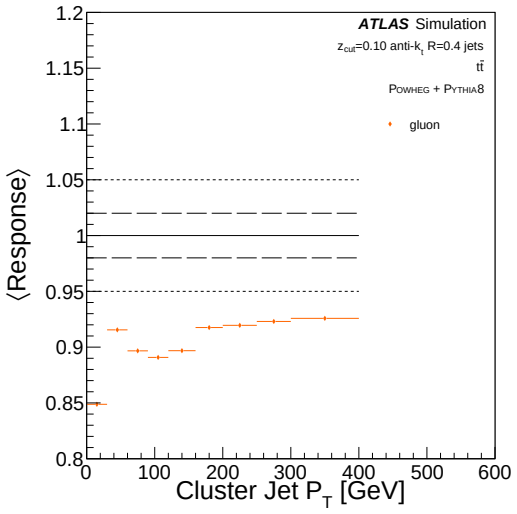


(f)

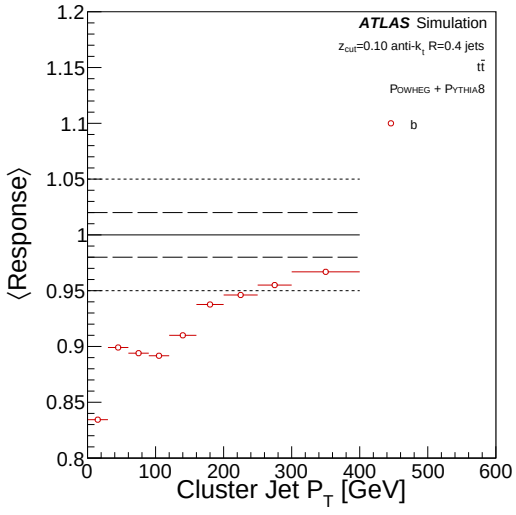
Figure F.5: Mean Jet Response of ungroomed uds-, (a), gluon-, (b), b-, (c), c-, (d), 1-pronged tau -, (e) and 3-pronged tau -jets, (f)



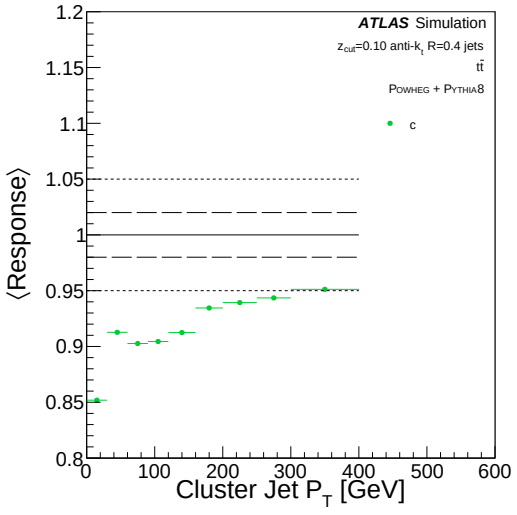
(a)



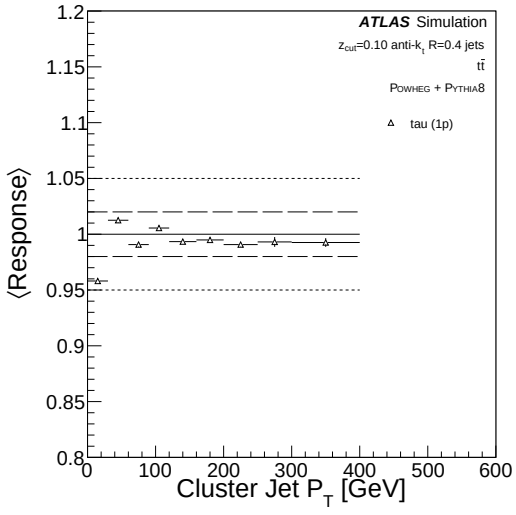
(b)



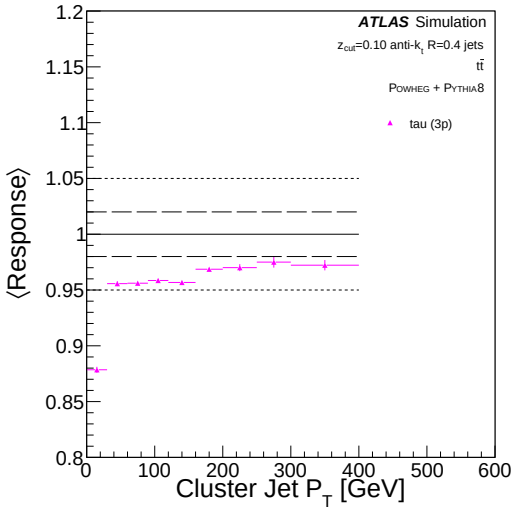
(c)



(d)

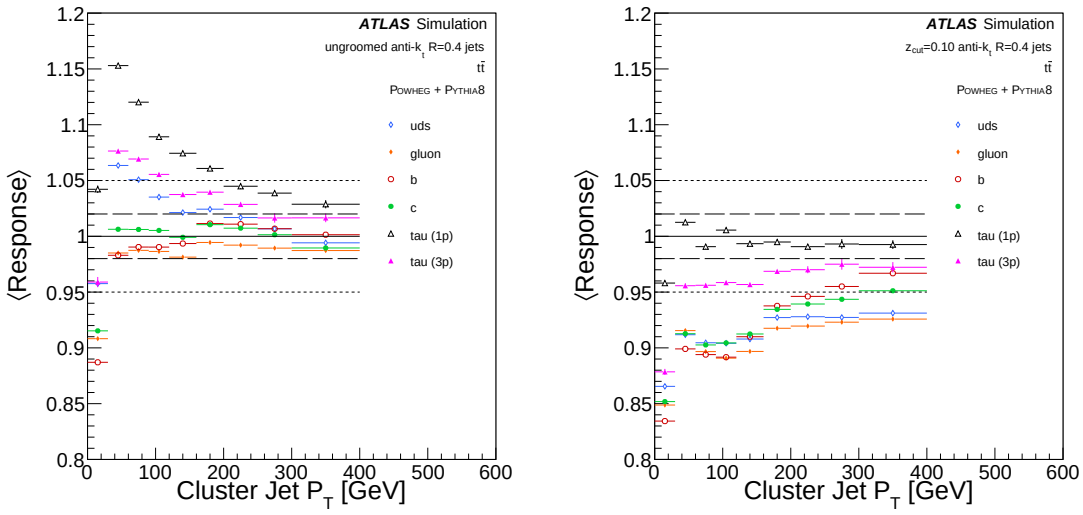


(e)



(f)

Figure F.6: Jet Response of SDb0z1 uds-, (a), gluon-, (b), b-, (c), c-, (d), 1-pronged tau -, (e) and 3-pronged tau -jets, (f)



(a) (b)

Figure F.7: Mean Jet Response of ungroomed, (a), and groomed, (b), jets.

GLOSSARY

4F four-flavour 56

5F five-flavour 56

AF2 AltFastII 59

ALICE A Large Ion Collider Experiment 29

AOD Analysis Objects Data 61

ATLAS A Toroidal LHC ApparatuS vi

BDT Boosted Decisions Tree 67

BR Branching Ratio 108

BSM Beyond Standard Model 23

C/A Cambridge/Aachen x

CERN European Organization for Nuclear Research ix

CKKW Catani-Krauss-Kuhn-Webber 58

CKM Cabibbo–Kobayashi–Maskawa 24

CMS Compact Muon Solenoid 22

CoM centre-of-mass viii

CPU Central Processing Unit 89

CSC Cathode Strip Chambers 49

CTP Central Trigger Processor 90

DGLAP Dokshitzer–Gribov–Lipatov–Altarelli–Parisi equations 56

DIS Deep Inelastic Scattering 54

DM Dark Matter 67

ECAL Electromagnetic Calorimeter ix

EF Event Filter 89

EM electromagnetic 3

RF radiofrequency 33

ROD Read Out Driver 91

ROS Read Out System 91

RoI Region of Interest 49

ROS Read-Out System 91

RPC Resistive-Plate Chambers 50

SCT SemiConductor Tracker 37

soft drop soft drop x

soft drop grooming soft drop grooming x

SDO Simulated Data Object 60

SM Standard Model iv

SPD Silicon Pixel Detector 41

SPS Super Proton Synchrotron 33

sPSB small Proton Synchrotron Booster 33

SUSY Supersymmetry 67

SV Secondary Vertex 63

TDAQ Trigger and Data Acquisition x

TGC Thin-Gap Chambers 50

TOTEM TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC 30

E_{T} Transverse Energy 90

TRT Transition Radiation Tracker 37

UE Underlying Event ix

UV Ultraviolet 57

VEV Vacuum Expectation Value ix

WP Working Point xii

BIBLIOGRAPHY

[1] V. Klochkov, *Anisotropic flow measurements at FAIR and SPS energies*, 2019. https://cds.cern.ch/record/2705196. Presented 20 Nov 2019.

[2] T. Economist, "Worth the wait." https://www.economist.com/graphic-detail/2012/07/04/worth-the-wait.

[3] L. Alvarez-Gaume and J. Ellis, *Eyes on a prize particle*, Nature Phys. **7** no. 1, (2011) 2–3. https://cds.cern.ch/record/1399903. Editorial Material.

[4] C. Grojean, *Higgs Physics,*. https://cds.cern.ch/record/2243593. 12 pages, contribution to the CERN in the Proceedings of the 2015 CERN-Latin-American School of High-Energy Physics, Ibarra, Ecuador, 4 - 17 March 2015.

[5] LHC Higgs cross-section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs cross-sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [hep-ph].

[6] ATLAS Collaboration Collaboration, *Combined measurements of Higgs boson production and decay using up to* 80 fb^{-1} *of proton-proton collision data at* $\sqrt{s}$ *= 13 TeV collected with the ATLAS experiment*, Phys. Rev. D **101** no. arXiv:1909.02845, (2019) 012002. 48 p. https://cds.cern.ch/record/2688596. 81 pages in total, author list starting page 65, 21 figures, 12 tables. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2018-57.

[7] ATLAS, "Atlas luminosity public results run 2." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[8] J. Boyd, *LHC Run-2 and Future Prospects*, Tech. Rep. arXiv:2001.04370, Jan, 2020. https://cds.cern.ch/record/2707815. 9 pages, 6 figure, proceedings from lecture given at CERN-JINR ESHEP 2019 summer school in St Petersburg, Russia. Submitted for publication in a CERN Yellow Report.

[9] C. Lefèvre, *The CERN accelerator complex,*. http://cds.cern.ch/record/1260465.

[10] ATLAS Collaboration Collaboration, *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997. http://cds.cern.ch/record/338080.

[11] The ATLAS Collaboration, "Atlas experiment - photos." http://atlasexperiment.org/photos/inner-detector-combined.html.

[12] The ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** no. 08, (2008) S08003. http://stacks.iop.org/1748-0221/3/i=08/a=S08003.

[13] S. Höche, *Introduction to parton-shower event generators*, 2014.

[14] L. R. F. Castillo, *The Search and Discovery of the Higgs boson.* 2053-2571. Morgan and Claypool Publishers, 2015. http://dx.doi.org/10.1088/978-1-6817-4078-2.

[15] J. Mahlstedt, *The ATLAS Hadronic Tau Trigger*, Journal of Physics: Conference Series **513** (2014) 012021.

[16] G. Aad et al., *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, The European Physical Journal C **77** no. 7, (2017). http://dx.doi.org/10.1140/epjc/s10052-017-5004-5.

[17] G. P. Salam, *Towards Jetography*, Eur. Phys. J. C **67** (2010) 637–686, arXiv:0906.1833 [hep-ph].

[18] N. Bartosik, *Diagram showing the common principle of identification of jets initiated by b-hadron decays*, 2016. http://bartosik.pp.ua/hep_sketches/btagging.

[19] M. e. a. Aaboud, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, Journal of High Energy Physics **2018** no. 8, (2018). http://dx.doi.org/10.1007/JHEP08(2018)089.

[20] A. H. e. a. Hoang, *Nonperturbative corrections to soft drop jet mass*, Journal of High Energy Physics **2019** no. 12, (2019). http://dx.doi.org/10.1007/JHEP12(2019)002.

[21] ATLAS Collaboration, A. Collaboration, "Approved plots daq." https://twiki.cern.ch/twiki/bin/view/atlaspublic/approvedplotsdaq#plots_figures_pictures.

[22] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 2016.

[23] A. J. Larkoski et al., *Soft drop*, Journal of High Energy Physics **2014** no. 5, (2014). http://dx.doi.org/10.1007/JHEP05(2014)146.

[24] The ATLAS Collaboration, *Observation of a New Particle in the Search for the Standard Model Higgs boson with the ATLAS Detector at the LHC*, Phys.Lett. **B716** (2012), arXiv:1207.7214 [hep-ex].

[25] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716** (2012), arXiv:1207.7235 [hep-ex].

[26] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588.

[27] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266. https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[28] Particle Data Group Collaboration, M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** no. 3, (2018) 030001.

[29] M. E. Peskin, D. V. Schroeder, *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[30] D0 Collaboration Collaboration, S. Abachi et al., *Observation of the Top Quark*, Phys. Rev. Lett. **74** (1995) 2632–2637. https://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[31] CDF Collaboration Collaboration, F. Abe et al., *Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74** (1995) 2626–2631. https://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[32] M. L. Perl et al., *Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation*, Phys. Rev. Lett. **35** (1975) 1489–1492. https://link.aps.org/doi/10.1103/PhysRevLett.35.1489.

[33] G. Arnison et al., *Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV*, Physics Letters B **122** no. 1, (1983) 103 – 116. http://www.sciencedirect.com/science/article/pii/0370269383911772.

[34] UA1 Collaboration, G. Arnison et al., *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, Phys. Lett. **B126** (1983) 398–410. [,7.55(1983)].

[35] L. Lederman and C. Hill, *Symmetry and the Beautiful Universe*. Prometheus Books, 2004. https://books.google.co.uk/books?id=X2QPAQAAMAAJ.

[36] C. Quigg, *The Electroweak Theory*, 2002.

[37] W. N. Cottingham and D. A. Greenwod, *An introduction to the standard model*. Cambridge, 1998.

[38] A. Pich, *The Standard Model of Electroweak Interactions*, arXiv:1201.0537 [hep-ph].

[39] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (1954) 191–195. https://link.aps.org/doi/10.1103/PhysRev.96.191.

[40] Y. Yokoi and S. Abe, *Derivation of Bose–Einstein and Fermi–Dirac statistics from quantum mechanics: gauge-theoretical structure*, Journal of Statistical Mechanics: Theory and Experiment **2018** no. 2, (2018) 023112. http://dx.doi.org/10.1088/1742-5468/aaa794.

[41] P. A. M. Dirac and R. H. Fowler, *The quantum theory of the electron*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **117** no. 778, (1928) 610–624, https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1928.0023. https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1928.0023.

[42] C. Rovelli, *Notes for a brief history of quantum gravity*, 2000.

[43] K. J. Barnes, *Group theory for the standard model of particle physics and beyond.* 6, 2010.

[44] S. Weinberg, *The quantum theory of fields. Vol. 2: Modern applications.* Cambridge University Press, 2013.

[45] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533. https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[46] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** no. 2, (1973) 652–657, https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf. https://doi.org/10.1143/PTP.49.652.

[47] A. Salam and J. Ward, *Electromagnetic and weak interactions*, Physics Letters **13** no. 2, (1964) 168 – 171. http://www.sciencedirect.com/science/article/pii/0031916364907115.

[48] A. Salam, *Gauge unification of fundamental forces*, Rev. Mod. Phys. **52** (1980) 525–538. https://link.aps.org/doi/10.1103/RevModPhys.52.525.

[49] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (1967) 1554–1561. http://link.aps.org/doi/10.1103/PhysRev.155.1554.

[50] T. Nakano and K. Nishijima, *Charge Independence for V-particles**, Progress of Theoretical Physics **10** no. 5, (1953) 581–582, https://academic.oup.com/ptp/article-pdf/10/5/581/5364926/10-5-581.pdf. https://doi.org/10.1143/PTP.10.581.

[51] M. Gell-Mann, *The interpretation of the new particles as displaced charge multiplets*, Nuovo Cim. **4** no. S2, (1956) 848–866.

[52] L. B. Okun, *Leptons and Quarks: Special Edition Commemorating the Discovery of the Higgs boson.* North-Holland, Amsterdam, Netherlands, 1982.

[53] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509. https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[54] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[55] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Advantages of the color octet gluon picture*, Physics Letters B **47** no. 4, (1973) 365 – 368. http://www.sciencedirect.com/science/article/pii/0370269373906254.

[56] H. Fritzsch and M. Gell-Mann, *Current algebra: Quarks and what else?*, eConf **C720906V2** (1972) 135–165, arXiv:hep-ph/0208010.

[57] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) 1343–1346. https://link.aps.org/doi/10.1103/PhysRevLett.30.1343.

[58] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349. https://link.aps.org/doi/10.1103/PhysRevLett.30.1346.

[59] A. Deur, S. J. Brodsky, and G. F. [de Téramond], *The QCD running coupling*, Progress in Particle and Nuclear Physics **90** (2016) 1 – 74. http://www.sciencedirect.com/science/article/pii/S0146641016300035.

[60] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*, Phys. Rev. **127** (1962) 965–970.

[61] ATLAS, CMS Collaboration, G. Aad et al., *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, JHEP **08** (2016) 045, arXiv:1606.02266 [hep-ex].

[62] CMS Collaboration Collaboration, A. M. Sirunyan et al., *Observation of Higgs boson Decay to Bottom Quarks*, Phys. Rev. Lett. **121** (2018) 121801. https://link.aps.org/doi/10.1103/PhysRevLett.121.121801.

[63] M. Aaboud et al., *Observation of $(H \to b\bar{b})$ decays and VH production with the ATLAS detector*, Physics Letters B **786** (2018) 59 – 86. http://www.sciencedirect.com/science/article/pii/S0370269318307056.

[64] D0 Collaboration Collaboration, "Useful diagrams of top signals and backgrounds." https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html.

[65] M. e. a. Aaboud, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Physics Letters B **784** (2018) 173–191. http://dx.doi.org/10.1016/j.physletb.2018.07.035.

[66] A. Collaboration, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, 2020.

[67] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production cross-section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004. https://link.aps.org/doi/10.1103/PhysRevLett.110.252004.

[68] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080.

[69] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054.

[70] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Physics Letters B **710** no. 4, (2012) 612 – 622. http://www.sciencedirect.com/science/article/pii/S0370269312002766.

[71] O. S. Brüuning and others, *LHC Design Report,*. https://cds.cern.ch/record/782076.

[72] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008).

[73] CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation **3** no. 08, (2008) S08004. http://stacks.iop.org/1748-0221/3/i=08/a=S08004.

[74] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST (2008).

[75] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3** no. 08, (2008) S08002. http://stacks.iop.org/1748-0221/3/i=08/a=S08002.

[76] TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** no. 08, (2008) S08007. http://stacks.iop.org/1748-0221/3/i=08/a=S08007.

[77] J. Pinfold, *The MoEDAL experiment at the LHC*, EPJ Web Conf. **145** (2017) 12002.

[78] LHCf Collaboration, *LHCf experiment: Technical Design Report*. Technical Design Report LHCf. CERN, Geneva, 2006. http://cds.cern.ch/record/926196.

[79] W. Herr and B. Muratori, *Concept of luminosity,*. https://cds.cern.ch/record/941318.

[80] CERN, "Cern timelines: The large hadron collider." https://timeline.web.cern.ch/timeline-header/93.

[81] CERN, "Longer term lhc schedule." https://lhc-commissioning.web.cern.ch/schedule/lhc-long-term.htm.

[82] J. M. Campbell, J. W. Huston, and W. J. Stirling, *Hard interactions of quarks and gluons: a primer for LHC physics*, Reports on Progress in Physics **70** no. 1, (2006) 89–193. https://doi.org/10.1088%2F0034-4885%2F70%2F1%2Fr02.

[83] Z. Marshall, *Simulation of Pile-up in the ATLAS Experiment*, Tech. Rep. ATL-SOFT-PROC-2013-030, CERN, Geneva, Oct, 2013. http://cds.cern.ch/record/1616394.

[84] J. Jowett et al., *The 2018 heavy-ion run of the LHC,*. https://cds.cern.ch/record/2693921.

[85] CERN, "Accelerators." https://home.cern/science/accelerators.

[86] K. Wille, *The physics of particle accelerators: an introduction*. Oxford Univ. Press, Oxford, 2000. http://cds.cern.ch/record/560708.

[87] B. R. Martin, *Nuclear and particle physics*. 11, 2009.

[88] A. Yamamoto, , et al., *The ATLAS central solenoid*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **584** no. 1, (2008) 53 – 74. http://www.sciencedirect.com/science/article/pii/S0168900207020414.

[89] The ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J**C70** , arXiv:1004.5293.

[90] The ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report,*. https://cds.cern.ch/record/1291633.

[91] ATLAS Collaboration Collaboration, M. Backhaus, *The upgraded Pixel Detector of the ATLAS Experiment for Run2 at the Large Hadron Collider*, Tech. Rep. ATL-INDET-PROC-2015-015, CERN, Geneva, Dec, 2015. http://cds.cern.ch/record/2110260.

[92] G. A. et al., *ATLAS pixel detector electronics and sensors*, Journal of Instrumentation **3** no. 07, (2008) P07007. http://stacks.iop.org/1748-0221/3/i=07/a=P07007.

[93] J. R. Pater, *The ATLAS SemiConductor Tracker operation and performance*, Journal of Instrumentation **7** no. 04, (2012) C04001–C04001. https://doi.org/10.1088%2F1748-0221%2F7%2F04%2Fc04001.

[94] A. Abdesselam et al., *The barrel modules of the ATLAS semiconductor tracker*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **568** no. 2, (2006) 642 – 671. http://www.sciencedirect.com/science/article/pii/S016890020601388X.

[95] A. Abdesselam et al., *The ATLAS semiconductor tracker end-cap module*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **575** no. 3, (2007) 353 – 389. http://www.sciencedirect.com/science/article/pii/S0168900207003270.

[96] A. Ahmad et al., *The Silicon microstrip sensors of the ATLAS semiconductor tracker*, Nucl. Instrum. Meth. A **578** (2007) 98–118.

[97] A. Bingül, *The ATLAS TRT and its Performance at LHC*, Journal of Physics: Conference Series **347** no. 1, (2012) 012025. http://iopscience.iop.org/article/10.1088/1742-6596/347/1/012025/meta.

[98] T. A. T. collaboration, E. Abat, et al., *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, Journal of Instrumentation **3** no. 02, (2008) P02013–P02013. https://doi.org/10.1088%2F1748-0221%2F3%2F02%2Fp02013.

[99] B. Dolgoshein, *Transition radiation detectors*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **326** no. 3, (1993) 434 – 469. http://www.sciencedirect.com/science/article/pii/016890029390846A.

[100] C. W. Fabjan and F. Gianotti, *Calorimetry for Particle Physics*, Rev. Mod. Phys. **75** no. CERN-EP-2003-075, (2003) 1243–1286. 96 p. http://cds.cern.ch/record/692252.

[101] The ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report,*. https://cds.cern.ch/record/331061.

[102] G. e. a. Aad, *Readiness of the ATLAS liquid argon calorimeter for LHC collisions*, The European Physical Journal C **70** no. 3, (2010) 723–753. http://dx.doi.org/10.1140/epjc/s10052-010-1354-y.

[103] ATLAS Collaboration Collaboration, *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996. https://cds.cern.ch/record/331062.

[104] G. e. a. Aad, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, The European Physical Journal C **70** no. 4, (2010) 1193–1236. http://dx.doi.org/10.1140/epjc/s10052-010-1508-y.

[105] A. Artamonov et al., *The ATLAS Forward Calorimeter*, Journal of Instrumentation **3** no. 02, (2008) P02010–P02010. https://doi.org/10.1088%2F1748-0221%2F3%2F02%2Fp02010.

[106] The ATLAS Collaboration, *ATLAS muon spectrometer: Technical design report,*. https://cds.cern.ch/record/331068.

[107] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche and others., *General-purpose event generators for LHC physics*, Phys.Rept. **504** (2011). http://arXiv.org/abs/1101.2599.

[108] J. Blümlein, *The theory of deeply inelastic scattering*, Progress in Particle and Nuclear Physics **69** (2013) 28–84. http://dx.doi.org/10.1016/j.ppnp.2012.09.006.

[109] P. Taels, *Quantum Chromodynamics at small Bjorken-x*, 2017.

[110] J. Gao, L. Harland-Lang, and J. Rojo, *The structure of the proton in the LHC precision era*, Physics Reports **742** (2018) 1–121. http://dx.doi.org/10.1016/j.physrep.2018.03.002.

[111] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438–450. [Yad. Fiz.15,781(1972)].

[112] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298–318.

[113] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641–653.

[114] F. Maltoni and T. Stelzer, *MadEvent: automatic event generation with MadGraph*, Journal of High Energy Physics **2003** no. 02, (2003) 027–027. http://dx.doi.org/10.1088/1126-6708/2003/02/027.

[115] Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics **2010** no. 6, (2010). http://dx.doi.org/10.1007/JHEP06(2010)043.

[116] Bernd A. Berg, *Introduction to Markov Chain Monte Carlo Simulations and their Statistical Analysis*, arXiv:cond-mat/0410490 [cond-mat].

[117] S. Catani, F. Krauss, R. Kuhn and B. Webber, *QCD Matrix Elements + Parton Showers*, JHEP 0111 (2001), arXiv:hep-ph/0109231 [hep-ph].

[118] M. L. Mangano, M. Moretti and R. Pittau, *Multijet Matrix Elements and Shower Evolution in Hadronic Collisions: $Wb\bar{b}$ + n Jets as a Case Study*, Nucl.Phys **B632** (2001), arXiv:0108069 [hep-ph].

[119] Torbjörn Sjöstrand and Stephen Mrenna and Peter Skands, *PYTHIA 6.4 physics and manual*, Journal of High Energy Physics **2006** no. 05, (2006) 026. http://stacks.iop.org/1126-6708/2006/i=05/a=026.

[120] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007.

[121] G. Corcella and others., *HERWIG 6: an event generator of hadron emission reactions with interfering gluons (including supersymmetric processes)*, J. High Energy Phys. **JHEP(2001)** (2001). http://iopscience.iop.org/article/10.1088/1126-6708/2001/01/010/meta.

[122] A. Kupco , *Cluster Hadronization in HERWIG 5.9*, arXiv:hep-ph/9906412 [hep-ph].

[123] B. Andersson, S. Mohanty and F. Soderberg, *Recent Developments in the Lund Model*, arXiv:hep-ph/0212122 [hep-ph].

[124] *ATLAS Pythia 8 tunes to 7 TeV datas*, Tech. Rep. ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov, 2014. https://cds.cern.ch/record/1966419.

[125] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Physical Review D **82** no. 7, (2010). http://dx.doi.org/10.1103/PhysRevD.82.074018.

[126] R. D. Field, *The Underlying Event in Hard Scattering Processes*, arXiv:hep-ph/0201192 [hep-ph].

[127] G. e. a. Aad, *The ATLAS Simulation Infrastructure*, The European Physical Journal C **70** no. 3, (2010) 823–874. http://dx.doi.org/10.1140/epjc/s10052-010-1429-9.

[128] GEANT4 Collaboration, *Geant4 a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003).

[129] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, J. Phys.: Conf. Ser **396** no. 022031, (2012).

[130] The ATLAS Collaboration, *ATLAS Computing : technical design report,*. https://cds.cern.ch/record/837738.

[131] T. Cornelissen and others., *Concepts, design and implementation of the ATLAS new tracking (NEWT)*,. http://cds.cern.ch/record/10201006.

[132] *The Optimization of ATLAS Track Reconstruction in Dense Environments*, Tech. Rep. ATL-PHYS-PUB-2015-006, CERN, Geneva, Mar, 2015. https://cds.cern.ch/record/2002609.

[133] M. e. a. Aaboud, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, The European Physical Journal C **77** no. 5, (2017). http://dx.doi.org/10.1140/epjc/s10052-017-4887-5.

[134] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** no. 2, (1987) 444 – 450. http://www.sciencedirect.com/science/article/pii/0168900287908874.

[135] *Vertex Reconstruction Performance of the ATLAS Detector at "$\sqrt{s}$ = 13 TeV"*, Tech. Rep. ATL-PHYS-PUB-2015-026, CERN, Geneva, Jul, 2015. https://cds.cern.ch/record/2037717.

[136] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, *Adaptive vertex fitting*, J. Phys. **G34** (2007) N343.

[137] ATLAS Collaboration, M. Aaboud et al., *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data*, JINST **14** no. 03, (2019) P03017, arXiv:1812.03848 [hep-ex].

[138] The ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. **C74** no. 10, (2014) 3071, arXiv:1407.5063 [hep-ex].

[139] The ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at* $\sqrt{s}$ = 13 TeV, Tech. Rep. ATL-PHYS-PUB-2016-015, CERN, Geneva, Aug, 2016. https://cds.cern.ch/record/2203514.

[140] The ATLAS Collaboration Collaboration, *Photon identification in 2015 ATLAS data*, Tech. Rep. ATL-PHYS-PUB-2016-014, CERN, Geneva, Aug, 2016. https://cds.cern.ch/record/2203125.

[141] ATLAS Collaboration Collaboration, *Measurements of the photon identification efficiency with the ATLAS detector using 4.9 fb*$^{-1}$ *of pp collision data collected in 2011*, Tech. Rep. ATLAS-CONF-2012-123, CERN, Geneva, Aug, 2012. https://cds.cern.ch/record/1473426.

[142] ATLAS Collaboration, B. Resende, *Muon identification algorithms in ATLAS*, PoS **EPS-HEP2009** (2009) 431.

[143] R. Nicolaidou et al., *Muon identification procedure for the ATLAS detector at the LHC using Muonboy reconstruction package and tests of its performance using cosmic rays and single beam data*, J. Phys. Conf. Ser. **219** (2010) 032052.

[144] The ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s}$ *=13 TeV*, Eur. Phys. J. C **76** no. 5, (2016) 292, arXiv:1603.05598 [hep-ex].

[145] ATLAS Collaboration, C. Galea, *Tau Lepton Reconstruction in ATLAS*, Nucl. Part. Phys. Proc. **287-288** (2017) 111–114.

[146] ATLAS Collaboration, G. Aad et al., *Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment*, Eur. Phys. J. C **76** no. 5, (2016) 295, arXiv:1512.05955 [hep-ex].

[147] S. Profumo, L. Giani, and O. F. Piattella, *An Introduction to Particle Dark Matter*, 2019.

[148] S. P. Martin, *A Supersymmetry Primer*, arXiv:hep-ph/9709356.

[149] The ATLAS Collaboration, *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at $\sqrt{s}$= 13 TeV*, Tech. Rep. ATL-PHYS-PUB-2015-027, CERN, Geneva, Jul, 2015. http://cds.cern.ch/record/2037904.

[150] A. Hrynevich, *ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2*, Journal of Instrumentation **12** no. 06, (2017) C06038–C06038. https://doi.org/10.1088%2F1748-0221%2F12%2F06%2Fc06038.

[151] G. P. Salam, *Towards jetography*, The European Physical Journal C **67** no. 3-4, (2010) 637–686. http://dx.doi.org/10.1140/epjc/s10052-010-1314-6.

[152] J. E. Huth et al., *Toward a standardization of jet definitions*,. https://cds.cern.ch/record/217490.

[153] G. Sterman and S. Weinberg, *Jets from Quantum Chromodynamics*, Phys. Rev. Lett. **39** (1977) 1436–1439. https://link.aps.org/doi/10.1103/PhysRevLett.39.1436.

[154] OPAL Collaboration, S. Kluth and A. Verbytskyi, *Measurements of jet rates with the anti-k_t and SISCone algorithms at LEP with the OPAL detector*, EPJ Web Conf. **141** (2017) 02003.

[155] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Physical Review D **48** no. 7, (1993) 3160–3166. http://dx.doi.org/10.1103/PhysRevD.48.3160.

[156] S. Catani et al., *Longitudinally-invariant k_T-clustering algorithms for hadron-hadron collisions*, Nuclear Physics B **406** no. 1, (1993) 187 – 224. http://www.sciencedirect.com/science/article/pii/055032139390166M.

[157] M. Wobisch and T. Wengler, *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*, 1999.

[158] Y. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *Better jet clustering algorithms*, Journal of High Energy Physics **1997** no. 08, (1997) 001–001. http://dx.doi.org/10.1088/1126-6708/1997/08/001.

[159] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-ktjet clustering algorithm*, Journal of High Energy Physics **2008** no. 04, (2008) 063–063. http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[160] L. e. a. Asquith, *Performance of Jet Algorithms in the ATLAS Detector*, Tech. Rep. ATL-PHYS-INT-2010-129, CERN, Geneva, Dec, 2010. https://cds.cern.ch/record/1311867.

[161] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, The European Physical Journal C **72** no. 3, (2012). http://dx.doi.org/10.1140/epjc/s10052-012-1896-2.

[162] M. e. a. Aaboud, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at s=13TeV with the ATLAS detector*, Physical Review D **96** no. 7, (2017). http://dx.doi.org/10.1103/PhysRevD.96.072002.

[163] ATLAS Collaboration, G. Aad et al., *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Eur. Phys. J. C **81** no. 8, (2021) 689, arXiv:2007.02645 [hep-ex].

[164] *Pile-up subtraction and suppression for jets in ATLAS*, Tech. Rep. ATLAS-CONF-2013-083, CERN, Geneva, Aug, 2013. https://cds.cern.ch/record/1570994.

[165] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Physics Letters B **659** no. 1-2, (2008) 119–126. http://dx.doi.org/10.1016/j.physletb.2007.09.077.

[166] G. e. a. Aad, *Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector*, Journal of High Energy Physics **2013** no. 9, (2013). http://dx.doi.org/10.1007/JHEP09(2013)076.

[167] *Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV*, Tech. Rep. ATLAS-CONF-2015-002, CERN, Geneva, Mar, 2015. http://cds.cern.ch/record/2001682.

[168] The ATLAS collaboration, *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s}$ = 8 TeV*, Tech. Rep. ATLAS-CONF-2015-017, CERN, Geneva, Apr, 2015. https://cds.cern.ch/record/2008678.

[169] *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2015-029, CERN, Geneva, Jul, 2015. https://cds.cern.ch/record/2037702.

[170] A. Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector*, 2015.

[171] The ATLAS Collaboration, *Tagging and suppression of pile-up jets with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2014-018, CERN, Geneva, May, 2014. https://cds.cern.ch/record/1700870.

[172] ATLAS Collaboration Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, tech. rep., CERN, Geneva, Jul, 2017. https://cds.cern.ch/record/2273281. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-013.

[173] ATLAS Collaboration Collaboration, S. Heer, *The secondary vertex finding algorithm with the ATLAS detector*, Tech. Rep. ATL-PHYS-PROC-2017-195, CERN, Geneva, Oct, 2017. https://cds.cern.ch/record/2287604.

[174] G. Piacquadio and C. Weiser, *A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS*, Journal of Physics: Conference Series **119** no. 3, (2008) 032032. https://doi.org/10.1088%2F1742-6596%2F119%2F3%2F032032.

[175] *Performance of b-jet identification in the ATLAS experiment*, Journal of Instrumentation **11** no. 04, (2016) P04008–P04008. http://dx.doi.org/10.1088/1748-0221/11/04/P04008.

[176] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*,.

[177] ATLAS Collaboration, I. Connelly, *Performance and calibration of b-tagging with the ATLAS experiment at LHC Run-2*, EPJ Web Conf. **164** (2017) 07025.

[178] ATLAS Collaboration, G. Aad et al., *Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV*, Eur. Phys. J. C **79** no. 10, (2019) 836, arXiv:1906.11005 [hep-ex].

[179] ATLAS Collaboration, M. Aaboud et al., *Measurement of jet-substructure observables in top quark, W boson and light jet production in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **08** (2019) 033, arXiv:1903.02942 [hep-ex].

[180] P. Loch, *Studies of Jet Shapes and Substructure with ATLAS*,. https://cds.cern.ch/record/1603345.

[181] J. M. Butterworth et al., *Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider*, Phys. Rev. Lett. **100** (2008) 242001. https://link.aps.org/doi/10.1103/PhysRevLett.100.242001.

[182] D. Krohn, J. Thaler, and L.-T. Wang, *Jet trimming*, Journal of High Energy Physics **2010** no. 2, (2010). http://dx.doi.org/10.1007/JHEP02(2010)084.

[183] M. D. Schwartz and H. X. Zhu, *Nonglobal logarithms at three loops, four loops, five loops, and beyond*, Phys. Rev. D **90** no. 6, (2014) 065004, arXiv:1403.4949 [hep-ph].

[184] L. Heinrich, *The ATLAS Trigger Core Configuration and Execution System in Light of the ATLAS Upgrade for LHC Run 2*, Tech. Rep. ATL-DAQ-PROC-2015-016. 8, CERN, Geneva, May, 2015. https://cds.cern.ch/record/2016643.

[185] L. Medina, R. Tomás, G. Arduini, and M. Napsuciale, *Assessment of the performance of High-Luminosity LHC operational scenarios: integrated luminosity and effective pile-up density*, Can. J. Phys. **97** no. 5, (2018) 498–508. 11 p. https://cds.cern.ch/record/2689342.

[186] ATLAS Collaboration Collaboration, R. Kopeliansky and A. Collaboration, *ATLAS Trigger and Data Acquisition Upgrades for the High Luminosity LHC,*. https://cds.cern.ch/record/2701977.

[187] ATLAS Collaboration Collaboration, *ATLAS level-1 trigger: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1998. https://cds.cern.ch/record/381429.

[188] ATLAS Collaboration Collaboration, P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 2003. https://cds.cern.ch/record/616089.

[189] S. e. a. Armstrong, *Architecture of the ATLAS High Level Trigger Event Selection Software*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **518** no. 1-2, (2004) 537–541. http://dx.doi.org/10.1016/j.nima.2003.11.079.

[190] ATLAS Collaboration, W. Panduro Vazquez, *The ATLAS Data Acquisition System: from Run 1 to Run 2*, Nucl. Part. Phys. Proc. **273-275** (2016) 939–944.

[191] M. zur Nedden, *The LHC Run 2 ATLAS trigger system: design, performance and plans*, Journal of Instrumentation **12** no. 03, (2017) C03024–C03024. https://doi.org/10.1088%2F1748-0221%2F12%2F03%2Fc03024.

[192] The ATLAS Collaboration, *Fast TracKer (FTK) Technical Design Report*, Tech. Rep. CERN-LHCC-2013-007. ATLAS-TDR-021, Jun, 2013. https://cds.cern.ch/record/1552953. ATLAS Fast Tracker Technical Design Report.

[193] ATLAS Collaboration, N. Ilic, *The ATLAS Fast Tracker and Tracking at the High-Luminosity LHC*, JINST **12** no. 02, (2017) C02052.

[194] ATLAS Collaboration Collaboration, *Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System*, Tech. Rep. CERN-LHCC-2013-018. ATLAS-TDR-023, Sep, 2013. https://cds.cern.ch/record/1602235. Final version presented to December 2013 LHCC.

[195] N. Anjos et al., *The ATLAS Jet Trigger for Initial LHC Run II*, Tech. Rep. ATL-DAQ-INT-2015-001, CERN, Geneva, Dec, 2015. https://cds.cern.ch/record/2116336.

[196] *Track Reconstruction Performance of the ATLAS Inner Detector at* $\sqrt{s} = 13$ *TeV*, tech. rep., CERN, Geneva, Jul, 2015. https://cds.cern.ch/record/2037683. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-018.

[197] G. Aad et al., *ATLAS data quality operations and performance for 2015–2018 data-taking*, Journal of Instrumentation **15** no. 04, (2020) P04003–P04003. http://dx.doi.org/10.1088/1748-0221/15/04/P04003.

[198] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv:1002.2581 [hep-ph].

[199] T. Gleisberg and others., *Event generation with SHERPA 1.1*, J. High Energy Phys. **JHEP(2009)** (2009). http://iopscience.iop.org/article/10.1088/1126-6708/2009/02/007/meta.

[200] G. e. a. Aad, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s}$=13 TeV with the ATLAS experiment*, Physics Letters B **810** (2020) 135797. http://dx.doi.org/10.1016/j.physletb.2020.135797.

[201] The ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Phys. Rev. **D96** no. 7, (2017) 072002, arXiv:1703.09665 [hep-ex].

[202] ATLAS Collaboration, I. Maznas, *FTK: The hardware Fast TracKer of the ATLAS experiment at CERN*, EPJ Web Conf. **137** (2017) 12001.

This thesis was typeset using the LaTeX typesetting system created by Leslie Lamport.
The body text size is set to 11 pt with *Utopia Regular* with *Fourier* font, part of TeX Live.
The bibliography was typeset using the ATLAS-paper style.